

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 17:59:34 ; Search time 7262.23 Seconds  
(without alignments)  
17784.976 Million cell updates/sec

Title: US-09-809-920-3

Perfect score: 6172  
Sequence: 1 GCGGGCTCCCTGAGCTGAA.....ATMAATGAGACTTAACTC 6172

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_com: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	6172	100.0	6172	9	AB007042	AB007042 Homo sapi
2	6172	100.0	6172	9	AF083551	AF083551 Homo sapi
3	6163.8	99.9	6189	9	AB011091	AB011091 Homo sapi
4	6124.8	99.2	6147	9	AF001690	AF001690 Homo sapi
5	5834.6	94.5	5854	9	AF029231	AF029231 Homo sapi
6	3217	52.1	3241	9	BC006363	BC006363 Homo sapi
7	3035	49.2	167343	6	AX332231	AX332231 Sequence
8	3035	49.2	167343	6	AX335067	AX335067 Sequence
9	3035	48.9	168194	2	HS096629	HS096629 Human chrom
10	3018.4	41.6	199912	2	AC040975	AC040975 Homo sapi
11	2567.4	39.9	3479	10	AC012411	AC012411 Homo sapi
12	2459.6	38.0	3479	10	AF083550	AF083550 Mus muscu
13	2347.4	31.98	3198	10	AB033367	AB033367 Rattus no
14	1077.8	17.5	236447	2	AC087099	AC087099 Mus muscu
15	621	10.1	64626	2	AC011641	AC011641 Homo sapi
16	596.4	9.7	598	6	AX062596	AX062596 Sequence
17	545.6	8.8	3622	3	AF132161	AF132161 Drosophi
18	545.6	8.8	106659	2	AC020331	AC020331 Drosophi
19	545.6	8.8	161172	3	AC007839	AC007839 Drosophi
20	545.6	8.8	287859	3	AE003797	AE003797 Drosophi
21	521.8	8.5	64626	2	AC011641	AC011641 Homo sapi
22	396	6.4	396	9	HS076188	HS076188 Homo sapi
23	256.2	4.2	2530	3	CE094835	CE094835 Caenorhab
24	256.2	4.2	2616	3	CE094835	CE094835 Caenorhab
25	244.6	4.0	23164	2	AC108250	AC108250 Rattus no
26	224.4	3.6	2860	10	MM067837	MM067837 Mus muscu
27	224.4	3.6	2862	10	BC006597	BC006597 Mus muscu
28	224.4	3.6	2864	10	MM072141	MM072141 Bos tauru
29	218.8	3.5	2884	4	AF089748	AF089748 Bos tauru
30	200.6	3.3	3069	9	BC010058	BC010058 Homo sapi
31	200.6	3.3	3175	9	HS062740	HS062740 Homo sapi
32	200.6	3.3	3320	9	HS064511	HS064511 Human herd
33	199	3.2	3003	6	AR168099	AR168099 Sequence
34	199	3.2	3003	9	HS072263	HS072263 Human mult
35	181.2	2.9	3077	10	AF252858	AF252858 Cricetulu
36	174.6	2.8	2562	10	BC004741	BC004741 Mus muscu
37	174.6	2.8	2708	10	MM078539	MM078539 Mus muscu
38	169.4	2.7	2769	9	BC001174	BC001174 Homo sapi
39	169.4	2.7	3183	9	S79639	S79639 EXT1-puclat
40	166.2	2.7	2709	3	AF145598	AF145598 Drosophi
41	166.2	2.7	3642	3	AF083889	AF083889 Drosophi
42	160.2	2.6	2965	6	AX306135	AX306135 Sequence
43	160.2	2.6	2965	10	MMEX01	MMEX01 M.musculus
44	151	2.4	2867	5	AF319538	AF319538 Xenopus 1
45	140.6	2.3	19442	3	CEK01G5	CEK01G5 Caenorhabd

## ALIGNMENTS

RESULT 1  
AB007042 6172 bp mRNA linear PRI 13-FEB-1999  
LOCUS AB007042 Homo sapiens EXTR1 mRNA, complete cds.  
DEFINITION AB007042  
ACCESSION AB007042.1 GI:2723390  
VERSION  
KEYWORDS  
SOURCE Homo sapiens testis and total fetus cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Direct Submission  
Submitted (06-SEP-1997) Yoshiyuki Saito, National Institute of  
Radiological Sciences, Genome Research Group, Anagawa 4-9-1, Inage,  
Chiba 263-8555, Japan (E-mail: C-saito@nirs.go.jp,  
Tel: 81-43-206-3135, Fax: 81-43-251-9818)  
2 (sites)  
REFERENCE  
Saito, T., Seki, N., Yamauchi, M., Tsuji, S., Hayashi, A., Kozuma, S. and  
Hori, T.  
Structure, chromosome location, and expression profile of EXTR1  
and EXTR2, new members of the multiple exostoses gene family

REFERENCE  
Saito, T., Seki, N., Yamauchi, M., Tsuji, S., Hayashi, A., Kozuma, S. and  
Hori, T.  
Structure, chromosome location, and expression profile of EXTR1  
and EXTR2, new members of the multiple exostoses gene family

JOURNAL Biochem. Biophys. Res. Commun. 243 (1), 61-66 (1998)  
MEDLINE 98139867  
COMMENT updated (19-Sep-1997).  
FEATURES Location/Qualifiers  
source 1. 6172  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8p"  
/tissue\_type="testis and total fetus"  
594..3353  
/gene="EXTR1"  
594..3353  
/gene="EXTR1"  
/codon\_start=1  
/product="EXTR1"  
/protein\_id="BAA24080.1"  
/db\_xref="GI:2723391"  
/translation="MTGYTMLRNGAGNGGOTCMLRWSNRILRWLSFTLFLVLPF  
LIAHYLTLLDEADAGKRIFGPRVGNELCEYKHYLDLCRISEYSEELLLEAKROE  
LNSEIATLNLKIEACKSIENAKODLQIKNYISOTESHYKELMONGPKSLPRL  
PEKIDAGLPKPKATGCRHLNCFDYSRCPPLSGPPVYVDSOQPFYSYLDPLVQAF  
QATARANYVTENDIACLYIVLSEMEPPVLRPAELKOLISLPHRTGHNHVI  
VVRKYLFTQGEKIESLRSLGEARFEBEMEDPADYDRIIATLKAODSKLD  
QVLEFETCKNPKPSLPTEWALCGEREDRELKISTFALLITPDPLVSSGCATR  
LFEALVGAIVVGLAEQVLYQDMLQNEALVYKRVTEVHEHLISLSDSLAM  
RROGFLMETEYFSTADSIENYVLIAMIRRIOPAPREPAALGEPHLEPHFEDVLP  
MADGDDIGVETPEPPYASPRYLKNTLTLYTDPRSNKCAPPEHLEPHFEDVLP  
SEAFPLSGSTFRPIGGAGSGSGKEFQALGQVPREQFTVMVLYEEVLEMSLER  
LNGLPYLNKVVVNNSPKLPSEDLPDIPVIMVIRKESLNLRFLPMNEIEAI  
LSDIDHALLRDELIMFGFRVREARDRIVGPGRHADIHQSLYNSNCSLSW  
LTGAAPFHRYAYLYSYVPOAIRDMVDEYINCEDIAMNFIIVSHITRPPKIVSRMT  
FRGCGCPALSHDSDSHFERHKCINFEKVKYGYMPLLYTOFRVDSVLEFKTLPHDKTK  
CFKPI"

BASE COUNT 1280 a 1681 c 1692 g 1519 t  
ORIGIN

Query Match 100.0%; Score 6172; DB 9; Length 6172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGGGTCCTGAGCTGGAGCCGAGAGCAAGCCCTGGAGTTCCTTTCAAGAGT 60  
DB 1 GCGGGGTCCTGAGCTGGAGCCGAGAGCAAGCCCTGGAGTTCCTTTCAAGAGT 60  
QY 61 CGTGGCTGAGAGTGTAAATGCTACACAGTCAAGAGAGAGAGGCTCTGAACACATGCG 120  
DB 61 CGTGGCTGAGAGTGTAAATGCTACACAGTCAAGAGAGAGAGGCTCTGAACACATGCG 120  
QY 121 CTGATTGTGGCAAGGCATCATTAAGAGCGGCAATTATTCTGTCCTTACCTATTACT 180  
DB 121 CTGATTGTGGCAAGGCATCATTAAGAGCGGCAATTATTCTGTCCTTACCTATTACT 180  
QY 181 GTATAACTGTGAATAGACACTATGATTTGTTGTCAGCAAAACCAAGAAAGAGC 240  
DB 181 GTATAACTGTGAATAGACACTATGATTTGTTGTCAGCAAAACCAAGAAAGAGC 240  
QY 241 TATGGCAATTTGAAAGAGTGTCTGATTTCCAGGGGTGTTTCTCGGGTTCATCATGAG 300  
DB 241 TATGGCAATTTGAAAGAGTGTCTGATTTCCAGGGGTGTTTCTCGGGTTCATCATGAG 300  
QY 301 TACCTCTCCCTTTCATCTCAGCAAGAAATGGGCACTTTTATGCTTTGATTAAGATTAA 360  
DB 301 TACCTCTCCCTTTCATCTCAGCAAGAAATGGGCACTTTTATGCTTTGATTAAGATTAA 360  
QY 361 GGAACATGTTCTTTGGTCAACAGCCAGAACTTAAATCTGCGGAATGGGGCAGAGACCA 420  
DB 361 GGAACATGTTCTTTGGTCAACAGCCAGAACTTAAATCTGCGGAATGGGGCAGAGACCA 420  
QY 421 TTTTCAGCTGAGCTGAGAGAAATGAATGTTCAATTTTATTTGTTGCTTCTGGGAGC 480  
DB 421 TTTTCAGCTGAGCTGAGAGAAATGAATGTTCAATTTTATTTGTTGCTTCTGGGAGC 480

DB 421 TTTTCAGCTGAGCTGAGAGAAATGAATGTTCAATTTTATTTGTTGCTTCTGGGAGC 480  
QY 481 ACACATACTTTCTGGAACGTCGTCAAGAGAGATGTTTGTGAAATGACAACC 540  
DB 481 ACACATACTTTCTGGAACGTCGTCAAGAGAGATGTTTGTGAAATGACAACC 540  
QY 541 ATGGTATGCGAGTGAACCCGAGCTGATCTGGGGGGCAGGCTGACAGAGACTATGACAG 600  
DB 541 ATGGTATGCGAGTGAACCCGAGCTGATCTGGGGGGCAGGCTGACAGAGACTATGACAG 600  
QY 601 GCTATACATGCTGCGGAATGAGGGGGCGGGGAAAGAGAGTCAAGCTGATCTGCGCT 660  
DB 601 GCTATACATGCTGCGGAATGAGGGGGCGGGGAAAGAGAGTCAAGCTGATCTGCGCT 660  
QY 661 GGTCCAAACCGCATCCGCTCTACGTCGCTCAGCTTCACGCTTTGTATCTGCTGCTCT 720  
DB 661 GGTCCAAACCGCATCCGCTCTACGTCGCTCAGCTTCACGCTTTGTATCTGCTGCTCT 720  
QY 721 TCCCGCTCATGCGCCCTATTACCTTACCTACCATCTGATGAGGCTGATGAGGCAAGC 780  
DB 721 TCCCGCTCATGCGCCCTATTACCTTACCTACCATCTGATGAGGCTGATGAGGCAAGC 780  
QY 781 GGAATTTTGTGTCGCCGGTGGGGAAAGAGCTGTCGAGGTGAAGCAAGCTGATCTGT 840  
DB 781 GGAATTTTGTGTCGCCGGTGGGGAAAGAGCTGTCGAGGTGAAGCAAGCTGATCTGT 840  
QY 841 GCCGATCCGGAGATCGGTGAGTGAAGAGCTCTCTGACGTGAGGCAAGCCCAAGAGC 900  
DB 841 GCCGATCCGGAGATCGGTGAGTGAAGAGCTCTCTGACGTGAGGCAAGCCCAAGAGC 900  
QY 901 TGAACAGGAGATTCGCAAGCTGAAATGCAATGGAAGCCCTGTAAAGAGACATTGAGA 960  
DB 901 TGAACAGGAGATTCGCAAGCTGAAATGCAATGGAAGCCCTGTAAAGAGACATTGAGA 960  
QY 961 AGCCAAAGAGAGCTGCTCCAGTCAAGAAATGTCATGACGACGAGCCGAGCTTCTCTCA 1020  
DB 961 AGCCAAAGAGAGCTGCTCCAGTCAAGAAATGTCATGACGACGAGCCGAGCTTCTCTCA 1020  
QY 1021 AGGAGCTATGCGCCCAAGAACAGCCCAAGCTGCTCCCTCCATCGAGTGTCCAGAGA 1080  
DB 1021 AGGAGCTATGCGCCCAAGAACAGCCCAAGCTGCTCCCTCCATCGAGTGTCCAGAGA 1080  
QY 1081 AGGAGATGCGCGGCTCTCCCTCCCGGAAGGCGACCTCGGCGCTACACAACCTCT 1140  
DB 1081 AGGAGATGCGCGGCTCTCCCTCCCGGAAGGCGACCTCGGCGCTACACAACCTCT 1140  
QY 1141 TTGATTTATTCGTTGCCCTCTCAACCTCTGCTTCCGCTCTACGTCATGACAGTGCAC 1200  
DB 1141 TTGATTTATTCGTTGCCCTCTCAACCTCTGCTTCCGCTCTACGTCATGACAGTGCAC 1200  
QY 1201 AGTTTGTCTTGGCAGCTACCTGATCCCTTGTGTCAGAGGCTTTTCAGGCGAGACAC 1260  
DB 1201 AGTTTGTCTTGGCAGCTACCTGATCCCTTGTGTCAGAGGCTTTTCAGGCGAGACAC 1260  
QY 1261 GAGCTAAGCTTTATGTTACAAATAAGGACATGCGCTGCTTACGTATTAAGTGG 1320  
DB 1261 GAGCTAAGCTTTATGTTACAAATAAGGACATGCGCTGCTTACGTATTAAGTGG 1320  
QY 1321 GAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGAGCTGAGAGAGAGTGTATTC 1380  
DB 1321 GAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGAGCTGAGAGAGAGTGTATTC 1380  
QY 1381 TGCCACACTGGCGAGCGAGATGACACAACATGTCATCATCAATCTGCAAGTAAGTAC 1440  
DB 1381 TGCCACACTGGCGAGCGAGATGACACAACATGTCATCATCAATCTGCAAGTAAGTAC 1440  
QY 1441 ATACACAGAACTTCTTATTAAGCTCACTGACTGCGCTGCGATGCTGCGGCAAGTCACT 1500  
DB 1441 ATACACAGAACTTCTTATTAAGCTCACTGACTGCGCTGCGATGCTGCGGCAAGTCACT 1500  
QY 1501 TCTACACTGTCAGTACAGAGCTGGCTTGTGACTTGTGCTGATACCGCTGTGCTCATGCA 1560  
DB 1501 TCTACACTGTCAGTACAGAGCTGGCTTGTGACTTGTGCTGATACCGCTGTGCTCATGCA 1560

Qy	1561	TGTCGAGCCCACTTCAAGAAATCCACACAGCTGCCGGTGAAGCGGAATATCTCT	1620
Db	1561	TT	1620
Qy	1561	TGTCGAGCCCACTTCAAGAAATCCACACAGCTGCCGGTGAAGCGGAATATCTCT	1620
Db	1561	TT	1620
Qy	1621	TCACGCTCCAGGGCGGAAGATTGAGTCTGTAGGCTTGTAGCCCTTCAGAGGGCCGCTCT	1680
Db	1621	TT	1680
Qy	1681	TCGAAAGAGAAATGAGAGGCGACCCCTCCGCGCACTACGATGACCGGATCATTCGCCACCC	1740
Db	1681	TT	1740
Qy	1741	TCGAAAGAGAAATGAGAGGCGACCCCTCCGCGCACTACGATGACCGGATCATTCGCCACCC	1800
Db	1741	TT	1800
Qy	1801	TCGAAAGAGAAATGAGAGGCGACCCCTCCGCGCACTACGATGACCGGATCATTCGCCACCC	1860
Db	1801	TT	1860
Qy	1861	AAATGCTGAAGCTCTCCACCTTCGCCCTCATCTTAACCCCGGGGACCCCTCGCTGGTTA	1920
Db	1861	TT	1920
Qy	1921	TTTTCTCTGGGTGTGCAACACGCGCTCTTGGAAAGCCCTGGAAAGTCGGTCCGCTCCGGTGG	1980
Db	1921	TT	1980
Qy	1981	TGCTGGGGGAGACAGTTCACAGCTCCCTACCAGAGACATGCTGCAGTGGAAAGAGCGGCC	2040
Db	1981	TT	2040
Qy	2041	TGCTGGGGGAGACAGTTCACAGCTCCCTACCAGAGACATGCTGCAGTGGAAAGAGCGGCC	2100
Db	2041	TT	2100
Qy	2101	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2160
Db	2101	TT	2160
Qy	2161	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2220
Db	2161	TT	2220
Qy	2221	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2280
Db	2221	TT	2280
Qy	2281	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2340
Db	2281	TT	2340
Qy	2341	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2400
Db	2341	TT	2400
Qy	2401	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2460
Db	2401	TT	2460
Qy	2461	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2520
Db	2461	TT	2520
Qy	2521	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2580
Db	2521	TT	2580
Qy	2581	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2640
Db	2581	TT	2640
Qy	2641	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2700
Db	2641	TT	2700
Qy	2701	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2760
Db	2701	TT	2760
Qy	2761	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2820
Db	2761	TT	2820
Qy	2821	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2880
Db	2821	TT	2880
Qy	2881	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	2940
Db	2881	TT	2940
Qy	2941	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3000
Db	2941	TT	3000
Qy	3001	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3060
Db	3001	TT	3060
Qy	3061	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3120
Db	3061	TT	3120
Qy	3121	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3180
Db	3121	TT	3180
Qy	3181	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3240
Db	3181	TT	3240
Qy	3241	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3300
Db	3241	TT	3300
Qy	3301	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3360
Db	3301	TT	3360
Qy	3361	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3420
Db	3361	TT	3420
Qy	3421	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3480
Db	3421	TT	3480
Qy	3481	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3540
Db	3481	TT	3540
Qy	3541	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3600
Db	3541	TT	3600
Qy	3601	GTGACCTCTGCTATAGAGGGGCGGAGCGGCTTCTCTGGGAGACTTACTTCTCCACCTG	3660
Db	3601	TTTT	

QY	2641	TGATGGCCCTCCCTTACCTTAACAAGGTGTTGTTGGAAATTCCTCCCAAGCTGCAT	2700
Db	2641	TGAATGGCCCTCCCTTACCTTAACAAGGTGTTGTTGGAAATTCCTCCCAAGCTGCAT	2700
QY	2701	CAGAGGACCTTCGTGGCCCTGACATTGGGGTTCCCATCATGSGTGGCTGCTGAGAA	2760
Db	2701	CAGAGGACCTTCGTGGCCCTGACATTGGGGTTCCCATCATGSGTGGCTGCTGAGAA	2760
QY	2761	ACAGTTGGAACAACCGATTCTTACCCCTGGAAATGAAATTGAGACAGAGGCCATCTGTCCA	2820
Db	2761	ACAGTTGGAACAACCGATTCTTACCCCTGGAAATGAAATTGAGACAGAGGCCATCTGTCCA	2820
QY	2821	TTGATGACATGCTCACTCTCCGCATGAGAAATCATTTTGGGTTCCGGGTGTGGAGAG	2880
Db	2821	TTGATGACATGCTCACTCTCCGCATGAGAAATCATTTTGGGTTCCGGGTGTGGAGAG	2880
QY	2881	AAGCTGGGAGCCGATGTTGGGCTTCCTCGGCCGTTACACGACATGGGACATCCGCCATC	2940
Db	2881	AAGCTGGGAGCCGATGTTGGGCTTCCTCGGCCGTTACACGACATGGGACATCCGCCATC	2940
QY	2941	AGTCTGGCTCTTACAACTCCAACTACTCTGTGAGCTGTCTCATGGTGTGACAGGTGCTG	3000
Db	2941	AGTCTGGCTCTTACAACTCCAACTACTCTGTGAGCTGTCTCATGGTGTGACAGGTGCTG	3000
QY	3001	CCCTCTTTCACAAATATTATATGCTTACCTGATTTCTTATGTATGTCGCCAGGCCATCCGGG	3060
Db	3001	CCCTCTTTCACAAATATTATATGCTTACCTGATTTCTTATGTATGTCGCCAGGCCATCCGGG	3060
QY	3061	ACATGTTGGAGTAATACATCACTGAGGAGACATTGCCATGAACTCCGTGCTCCCA	3120
Db	3061	ACATGTTGGAGTAATACATCACTGAGGAGACATTGCCATGAACTCCGTGCTCCCA	3120
QY	3121	TCACCTGGGAAGCCCCCATCAAGGTACCTCAACGATGAGCATTTCCGATGCCAGATGCC	3180
Db	3121	TCACCTGGGAAGCCCCCATCAAGGTACCTCAACGATGAGCATTTCCGATGCCAGATGCC	3180
QY	3181	CTCAGGCCCTGTCTCATGATGACTCCCACTTCAGAGCGGCACAAATGTCATCAACTCT	3240
Db	3181	CTCAGGCCCTGTCTCATGATGACTCCCACTTCAGAGCGGCACAAATGTCATCAACTCT	3240
QY	3241	TCGTGAAGGTGTACGGTACATCCCTCTCTGACAGGAGTTCAAGGGGGAATCTGTGTC	3300
Db	3241	TCGTGAAGGTGTACGGTACATCCCTCTCTGACAGGAGTTCAAGGGGGAATCTGTGTGC	3300
QY	3301	TCCTTCAGACACGGCTCCGCCATGACAAAGCACAAAGTCTCAAGTTCATCTGAGGGCAGC	3360
Db	3301	TCCTTCAGACACGGCTCCGCCATGACAAAGCACAAAGTCTCAAGTTCATCTGAGGGCAGC	3360
QY	3361	GCACGGTCTGGGGAAGAAGATGAGCAGAGGGAAGAAATGGCTCCCAAGTTCCTAGGCA	3420
Db	3361	GCACGGTCTGGGGAAGAAGATGAGCAGAGGGAAGAAATGGCTCCCAAGTTCCTAGGCA	3420
QY	3421	TTGCAAGACCTTGGGCACATCTGCTGTGGTGGTGGCCAGAGCCTCTGCTGGAAGGGCAG	3480
Db	3421	TTGCAAGACCTTGGGCACATCTGCTGTGGTGGTGGCCAGAGCCTCTGCTGGAAGGGCAG	3480
QY	3481	CAGAGGAGTGTGAAGGAAACCGCTGCTTATCTTAAATCAAGCCACACTGGGCTGTGAG	3540
Db	3481	CAGAGGAGTGTGAAGGAAACCGCTGCTTATCTTAAATCAAGCCACACTGGGCTGTGAG	3540
QY	3541	CCCTGGGGGAGATCCCGGGGTTCCCAACAGAGGAGCATGATGATAGCTTACACTGAGG	3600
Db	3541	CCCTGGGGGAGATCCCGGGGTTCCCAACAGAGGAGCATGATGATAGCTTACACTGAGG	3600
QY	3601	ACTGTGGGACTGTGAGAGTCACTACACCGTTTCTGACGCCACAGACAGCTGTGTTCTGTG	3660
Db	3601	ACTGTGGGACTGTGAGAGTCACTACACCGTTTCTGACGCCACAGACAGCTGTGTTCTGTG	3660
QY	3661	GTTTTTACATTCAATTAACAATCTTATATGATATTAAAAAGGAAAGTTTCAGATTGTGCC	3720
Db	3661	GTTTTTACATTCAATTAACAATCTTATATGATATTAAAAAGGAAAGTTTCAGATTGTGCC	3720
QY	3721	ATTCAGGCTATTATTATATATATGTGTGTATATATAATATGACACACTGTGCATCA	3780

Db	3721	ATTCAAGGCTATATTTATATATATGCTGTATATATAAATACATGACACACTTGATACA	3780
QY	3781	TATATATTTTTGGCTGGGGGAGTGTGACTTTTGCCTTTCTAAAGGAGGACGGGACGGC	3840
Db	3781	TATATATTTTTGGCTGGGGGAGTGTGAGTTTGGCTTTCTAAAGGAGGACGGGACGGC	3840
QY	3841	TCCTTTGTCCTATTTCTGAGGCAATGGGTCTGGCCCTTGTCTACTGCTTATCCCTTA	3900
Db	3841	TCCTTTGTCCTATTTCTGAGGCAATGGGTCTGGCCCTTGTCTACTGCTTATCCCTTA	3900
QY	3901	AAGATCATCTCCATCTCCACCGAGGCAATCGTGTGAGCAACCAAGAAAGGATGAAT	3960
Db	3901	AAGATCATCTCCATCTCCACCGAGGCAATCGTGTGAGCAACCAAGAAAGGATGAAT	3960
QY	3961	TGGCCCTCTGGGGGCGTGAGCAAGGTCCTCCCTTAACCCCTTCTGTGTGGCAGACACAA	4020
Db	3961	TGGCCCTCTGGGGGCGTGAGCAAGGTCCTCCCTTAACCCCTTCTGTGTGGCAGACACAA	4020
QY	4021	CCTGTAACTCACTTCTTCTCCACGTGAATCCCTGGGAGCGCTGACCCCTGTGGGCTGT	4080
Db	4021	CCTGTAACTCACTTCTTCTCCACGTGAATCCCTGGGAGCGCTGACCCCTGTGGGCTGT	4080
QY	4081	TCAGCTTCTCTGTGTGGGGCAGCAGATTTTGTAGAGATTATCTTTAGGCCAGGCTTGCC	4140
Db	4081	TCAGCTTCTCTGTGTGGGGCAGCAGATTTTGTAGAGATTATCTTTAGGCCAGGCTTGCC	4140
QY	4141	TCGCTACTTATCCCTGCTCTCCACATTTCTCTCTGTGTGAGAGAGATGAGGAAGAAG	4200
Db	4141	TCGCTACTTATCCCTGCTCTCCACATTTCTCTCTGTGTGAGAGAGATGAGGAAGAAG	4200
QY	4201	AGTAGAAGAAATAGGGGGCTGGAAGACCCACCTCCACAGATGGCTTTTCTATCCTGCTTT	4260
Db	4201	AGTAGAAGAAATAGGGGGCTGGAAGACCCACCTCCACAGATGGCTTTTCTATCCTGCTTT	4260
QY	4261	CTGTGTGAACAACAGTCTCTGTGGGCCACAGGCGTTCTGAAAGGCTCTTTTGGATTGG	4320
Db	4261	CTGTGTGAACAACAGTCTCTGTGGGCCACAGGCGTTCTGAAAGGCTCTTTTGGATTGG	4320
QY	4321	ACAGAGATCAGACAGCGTGACATCTGCTGTGTAAGTGGTTTGGACGTCAGACCTCC	4380
Db	4321	ACAGAGATCAGACAGCGTGACATCTGCTGTGTAAGTGGTTTGGACGTCAGACCTCC	4380
QY	4381	TCATCCAGTGTAGAGACCAAGCCAGTCTCTTGAGAGAACCCACCCGGCTGGCCGGGAAGT	4440
Db	4381	TCATCCAGTGTAGAGACCAAGCCAGTCTCTTGAGAGAACCCACCCGGCTGGCCGGGAAGT	4440
QY	4441	TTTACAGCAAGGCGCTGCTGGGTAATTCCTTGGTAATTCACCTTCCCCCGCT	4500
Db	4441	TTTACAGCAAGGCGCTGCTGGGTAATTCCTTGGTAATTCACCTTCCCCCGCT	4500
QY	4501	CTGTCTGGAGCCCATCTGTGTATCTGTGTGTTTTTGGACCCCTAATGTACGCTGGCT	4560
Db	4501	CTGTCTGGAGCCCATCTGTGTATCTGTGTGTTTTTGGACCCCTAATGTACGCTGGCT	4560
QY	4561	GTAGAGCTCCCGAGGTTTGGTAATGTGCTAAGAACATGTGGAGGCTGTGATTTGCTGTGA	4620
Db	4561	GTAGAGCTCCCGAGGTTTGGTAATGTGCTAAGAACATGTGGAGGCTGTGATTTGCTGTGA	4620
QY	4621	AGCTACATCCAGCCTTGGAAATCTAAGCGGCAATTCACAAACCGAGTTTACCACTTCCACT	4680
Db	4621	AGCTACATCCAGCCTTGGAAATCTAAGCGGCAATTCACAAACCGAGTTTACCACTTCCACT	4680
QY	4681	CCCTCCTTAGGATTTCTGTCCCTGGGCTGAATAGTAATAGCTAATTTTGGGTACG	4740
Db	4681	CCCTCCTTAGGATTTCTGTCCCTGGGCTGAATAGTAATAGCTAATTTTGGGTACG	4740
QY	4741	GTGGCAGTAGGGAACCTAGAGAGGCTGTGAATGTCATTTGTCAAGGATTTAGCCCATGAC	4800
Db	4741	GTGGCAGTAGGGAACCTAGAGAGGCTGTGAATGTCATTTGTCAAGGATTTAGCCCATGAC	4800
QY	4801	GTGTTTCTTGAACCTACTTCTCTGGAAGTGAAGTGTGACTTGAAGTTTCTAGCACTG	4860
Db	4801	GTGTTTCTTGAACCTACTTCTCTGGAAGTGAAGTGTGACTTGAAGTTTCTAGCACTG	4860

Db	4801	GTGTTTCTTGAAACCCCTACCTTTCTGTGAAGTGGAGTTGAGCTCTGGAGATTCTTCTAGCAACTG	4866
QY	4861	AACAAAGCTCAGAGTTTGTCTGTGCATCATCATGCGTTAAGCCAGTTCCGCTTCCCTA	4920
Db	4861	AACAAAGCTCAGAGTTTGTCTGTGCATCATCATGCGTTAAGCCAGTTCCGCTTCCCTA	4920
QY	4921	GACCTTGGCATCCTGTGCTTCTATTCTTCTGGAATACGTTCTCCTTGACAGCCGTAC	4980
Db	4921	GACCTTGGCATCCTGTGCTTCTATTCTTCTGGAATACGTTCTCCTTGACAGCCGTAC	4980
QY	4981	ACGTGGGCTCTTTCAGAGACTGTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCAC	5040
Db	4981	ACGTGGGCTCTTTCAGAGACTGTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCAC	5040
QY	5041	CTGTAGGGCTAGCTGGCTTAAGGGAATCGTCCCATTTGGCAACCGGACCCGGCCGCG	5100
Db	5041	CTGTAGGGCTAGCTGGCTTAAGGGAATCGTCCCATTTGGCAACCGGACCCGGCCGCG	5100
QY	5101	CCAGACTGTGTTTCCAAAGGTTCCCGCCGCCCAACCCAGACATCAGCTGTAGCTCCC	5160
Db	5101	CCAGACTGTGTTTCCAAAGGTTCCCGCCGCCCAACCCAGACATCAGCTGTAGCTCCC	5160
QY	5161	TGCTGAGGCACTGAGGTATATTTCTCCAGAGAGTGGGGTCAAGAGCCCTTCTCAGAACTT	5220
Db	5161	TGCTGAGGCACTGAGGTATATTTCTCCAGAGAGTGGGGTCAAGAGCCCTTCTCAGAACTT	5220
QY	5221	TCTAGTTGCCCTACCTGACTCCTGACTGATATATCTTTTACAGTACGCTTCCCT	5280
Db	5221	TCTAGTTGCCCTACCTGACTCCTGACTGATATATCTTTTACAGTACGCTTCCCT	5280
QY	5281	CGGGGAGCCAAAGAGTGTGTGTGGCGCTATATTTGTGCTCTATTATCTGCTTTC	5340
Db	5281	CGGGGAGCCAAAGAGTGTGTGTGGCGCTATATTTGTGCTCTATTATCTGCTTTC	5340
QY	5341	TTTTTAATGTAGAGAACTCAATATCTGACTCATGTGGGACTCGGTGAGCCGGGCGCTTG	5400
Db	5341	TTTTTAATGTAGAGAACTCAATATCTGACTCATGTGGGACTCGGTGAGCCGGGCGCTTG	5400
QY	5401	TGTGTGTGGGACCCCTTTTACGGGAGCTCAGTGAAGCTGGGGCGGTGTGTGTGGAGCCA	5460
Db	5401	TGTGTGTGGGACCCCTTTTACGGGAGCTCAGTGAAGCTGGGGCGGTGTGTGTGGAGCCA	5460
QY	5461	GGGCGCTCTCCCTTTAATGTGAAGCCAGTGTGTGGGCCCGGAATGTCACTGTGGATCTAAG	5520
Db	5461	GGGCGCTCTCCCTTTAATGTGAAGCCAGTGTGTGGGCCCGGAATGTCACTGTGGATCTAAG	5520
QY	5521	AAGGCGTGAAGTGTGTGACACCAAAACATATCCGAGGAGGGGCTGTGTGTGCGCGGTTC	5580
Db	5521	AAGGCGTGAAGTGTGTGACACCAAAACATATCCGAGGAGGGGCTGTGTGTGCGCGGTTC	5580
QY	5581	CAACAAGACAGACCCCTCTTGAACCTGAAAGGAMACACTGGCTTGAAGAGACTGCAGACAGG	5640
Db	5581	CAACAAGACAGACCCCTCTTGAACCTGAAAGGAMACACTGGCTTGAAGAGACTGCAGACAGG	5640
QY	5641	CTCTGAGGGGACAGCGCTCTCTAGGAGAGCAGTAAAGTGGCCACTGTGCTACTGTGTCA	5700
Db	5641	CTCTGAGGGGACAGCGCTCTCTAGGAGAGCAGTAAAGTGGCCACTGTGCTACTGTGTCA	5700
QY	5701	GGTGTCTTTCACACAGGGGAAAAGCCGCAACCTGTGACTCGTTTGAATGGGAAAAGCCGCG	5760
Db	5701	GGTGTCTTTCACACAGGGGAAAAGCCGCAACCTGTGACTCGTTTGAATGGGAAAAGCCGCG	5760
QY	5761	CCACAGACCCCGGGTCTCTCTTGGCTGTGTGAGGACCGCCCTGCGCACCTGTGTCTGGCT	5820
Db	5761	CCACAGACCCCGGGTCTCTCTTGGCTGTGTGAGGACCGCCCTGCGCACCTGTGTCTGGCT	5820
QY	5821	CGCAGAGGTGAGAGAGCGCTGCTCTCTGAGGTGGCCGGCTTGGCTCGCGTTTGGGCTTG	5880
Db	5821	CGCAGAGGTGAGAGAGCGCTGCTCTCTGAGGTGGCCGGCTTGGCTCGCGTTTGGGCTTG	5880
QY	5881	TCTTACCAATAACACCGTCCAGGCGCTCTGAGAGGCACTGTGAGAGCGGTGCTCCTGGGCA	5940
Db	5881	TCTTACCAATAACACCGTCCAGGCGCTCTGAGAGGCACTGTGAGAGCGGTGCTCCTGGGCA	5940



RESULT 2  
AF083551  
LOCUS Homo sapiens exostosin (EXTL3) mRNA, complete cds.  
DEFINITION AF083551  
ACCESSION AF083551.1 GI:5326770  
VERSION  
KEYWORDS  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 6172)  
AUTHORS Sato, T.  
TITLE The human EXTL3/EXTL3 gene at 9p11-p12, the third breast cancer susceptibility gene locus, is not mutated in breast and various cancers  
FEATURES  
SOURCE Unpublished  
2 (bases 1 to 6172)  
Sato, T.  
DIRECT SUBMISSION  
SUBMITTED 11-AUG-1998 Otolaryngology/Pathology, Columbia University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA  
LOCATION/Qualifiers  
1..6172  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="9p11-p12"  
1..6172  
/gene="EXTL3"  
/note="EXTL3; EXTL3; candidate gene for breast cancer susceptibility BRCA3"  
594..3353  
/gene="EXTL3"  
/codon\_start=1  
/product="exostosin"  
/db\_xref="GI:5326771"  
/db\_xref="AA02041.1"  
/translation="MTGYTLNRNGAGNGGOTCMIRMSNRIILTWLSFTLVILVEEP  
LIAHYVLTLDDEADKRIEFGVNELEKIVLDLCRIEVSSEELQLEKROE  
LNSFIKLTIEACKSINAKODLLOLVISQHSKELKLNRLPNNEITEAI  
PERDDALPPKATRGCRNLNCDISKCPITSGPRVYVSDOVSQSDPLVAKOF  
QATIRANVYITENDIACLVYLVGMEQEVLPALAEQQLSLPHWRDGHVIT  
NLRSKSTONLNLVSTGRAMVASTETVQYRFGDLVSPVLAHASENFMEIPQ  
VPAVKRYLFTFOGKIESLSLOEARSFEEMGDPADYDRIIATLAVODSKD  
QVLEVFKNQPKPSLPTLEWALGSEDRLELKLSTFALITPGDRLYISSCATR  
LFEALVGAVPVVGEOVLPIYDMLQWNEALVVPKPRVTEVAFILRSLSDDLAM  
RROGRFLMTEVFSTADISFNTVLAITRTIOTPAAPREAAAEIPRISKAQTDN  
MADNGDDLDGPEVEPPYASPRIRNTFLVTDFTYSWNCAPGFHLPPTPDPVLP  
SEAKFLSGTGFRPIGGAGSGKEFGAALGAVPRQFTVMTYREEVLMNSLER  
LNGILPYLNKVVVWNSPKLPSSEDLMPDIDGVPIMVVTENKSLNRLPNNEITEAI  
LSIDDDAHLRHDEIMFGFVYMRARDRIIVGPGRYHAMDIPHOSMLYNSYSELAW  
LTGAAPFKYVAYLYSVMPQAIIRDVDEYINCEDIAMNFIYITRKPKIYTSMT  
FRGCGPQALSHSDSHFERHKCNFVVKYGVYPLLYTQPRVDSVLFKRLRPHDKK  
CFKRT"

BASE COUNT 1280 a 1681 c 1692 g 1519 t  
ORIGIN  
Query Match 100.0%; Score 6172; DB 9; Length 6172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGGGTCCCTGAGCTGGAAAGCCGAGAGCAAGCCCTGAGAGTTCACTCTTCAAGAGT 60  
DB 1 GCGGGTCCCTGAGCTGGAAAGCCGAGAGCAAGCCCTGAGAGTTCACTCTTCAAGAGT 60  
QY 61 CGTGTCTGAGCTGTAATGCTACACAGTCAGAGAAAGAGGCTCTGAAACATATGC 120  
DB 61 CGTGTCTGAGCTGTAATGCTACACAGTCAGAGAAAGAGGCTCTGAAACATATGC 120  
QY 121 CTGATGTGGCAAGCATATGAAGCGGATTTATTTCTGTCTAATCACTTACT 180  
DB 121 CTGATGTGGCAAGCATATGAAGCGGATTTATTTCTGTCTAATCACTTACT 180  
QY 181 GTATACTGTAATAGACATATGATATTTGTCAGCAAAACAGAAAGAGC 240  
DB 181 GTATACTGTAATAGACATATGATATTTGTCAGCAAAACAGAAAGAGC 240  
QY 241 TATGGCATTTGAAAAAGCTGTGTGATTCAGGGTGTCTGGGTTTCATCATCAGG 300  
DB 241 TATGGCATTTGAAAAAGCTGTGTGATTCAGGGTGTCTGGGTTTCATCATCAGG 300  
QY 301 TACCTCCCTCTTCATCTGACGAAGAATGTGCACTTTATGTTTGAATAAGATTAA 360  
DB 301 TACCTCCCTCTTCATCTGACGAAGAATGTGCACTTTATGTTTGAATAAGATTAA 360  
QY 361 GGACATGTCTTGTGTCACACAGCAGCAAACTTAAATCTGCTGAATAGGTCAGAGCA 420  
DB 361 GGACATGTCTTGTGTCACACAGCAGCAAACTTAAATCTGCTGAATAGGTCAGAGCA 420  
QY 421 TTTACGCTGACGTCGAGAGAAATGAATGTCATTTATTTGTCCTGTCTGGGAGC 480  
DB 421 TTTACGCTGACGTCGAGAGAAATGAATGTCATTTATTTGTCCTGTCTGGGAGC 480  
QY 481 ACACACTCTTCTGGAAGAGTGTGAGGAACAGAGATGCTTTGGAATATGCAACCC 540  
DB 481 ACACACTCTTCTGGAAGAGTGTGAGGAACAGAGATGCTTTGGAATATGCAACCC 540  
QY 541 ATGGTATGGGAGTGAACCCGACGTGATCTGGGGGAGAGCTGACAGAGACTCATGACG 600  
DB 541 ATGGTATGGGAGTGAACCCGACGTGATCTGGGGGAGAGCTGACAGAGACTCATGACG 600  
QY 601 GCTATACCATCTCGCGAATGGGGGCGGGGAACGAGGTCAAGCCTGCATCTGGCGCT 660  
DB 601 GCTATACCATCTCGCGAATGGGGGCGGGGAACGAGGTCAAGCCTGCATCTGGCGCT 660  
QY 661 GGTCACACCGCATCGGCTCGTGCAGTCCAGCTTCAGCTTGTGATCTGCTGCTTCT 720  
DB 661 GGTCACACCGCATCGGCTCGTGCAGTCCAGCTTCAGCTTGTGATCTGCTGCTTCT 720  
QY 721 TCCCGCTCATCGCCACATATTAACCTCAACACTCTGATGAGAGCTGATGAGGACAGAC 780  
DB 721 TCCCGCTCATCGCCACATATTAACCTCAACACTCTGATGAGAGCTGATGAGGACAGAC 780  
QY 781 GGAATTTTGGTCCCGGGGAGGAGACGAGTGTGCGAGAGTGAACGACGTGCTGATCTGT 840  
DB 781 GGAATTTTGGTCCCGGGGAGGAGACGAGTGTGCGAGAGTGAACGACGTGCTGATCTGT 840  
QY 841 GCGCATCCCGGAGTGGTGATGAAGAGCTCGAGAGCTGAGAGGCAAGGCGCAAGAGC 900  
DB 841 GCGCATCCCGGAGTGGTGATGAAGAGCTCGAGAGCTGAGAGGCAAGGCGCAAGAGC 900  
QY 901 TGAACAGAGAGATGCGCAAGCTGATGAGATGAGAGCTGTAAGAGAGATTTGAGA 960  
DB 901 TGAACAGAGAGATGCGCAAGCTGATGAGATGAGAGCTGTAAGAGAGATTTGAGA 960  
QY 961 AGCGCAAGCAGACCTGCTCCAGCTCAAGAATGTATCATGACGACGAGCATTTCTTACA 1020

|||||  
Db 961 ACCGCAAGCAGACCTGCTCAGCTCAAGATGTCATCAGCAGACCGACATTTCTACA 1020  
QY 1021 AGGAGCTCATGAGCCAGAACACAGCTGTCCTGCTCCATCCGACTCTCCACAGA 1080  
Db 1021 AGGAGCTCATGAGCCAGAACACAGCTGTCCTGCTCCATCCGACTCTCCACAGA 1080  
QY 1081 AGGAGCATGCGGCGCTCCCTCCCGAAGGCACTGGGGCTGCGGCTACACATGCT 1140  
Db 1081 AGGAGCATGCGGCGCTCCCTCCCGAAGGCACTGGGGCTGCGGCTACACATGCT 1140  
QY 1141 TTGATTAATTCGTGTTGCCCTCTCAGCTTGCTCCGGTCTACGTCTATGACAGTAC 1200  
Db 1141 TTGATTAATTCGTGTTGCCCTCTCAGCTTGCTCCGGTCTACGTCTATGACAGTAC 1200  
QY 1201 AGTTTCTTTGGAGCTACCTGATCCCTTGCTGTCAGAGGCTTTTACAGGACAGCAC 1260  
Db 1201 AGTTTCTTTGGAGCTACCTGATCCCTTGCTGTCAGAGGCTTTTACAGGACAGCAC 1260  
QY 1261 GAGCTAACGTTTATGTTACAGAAATGCAGACATCGCTGCTTTACGTATCTAGTGG 1320  
Db 1261 GAGCTAACGTTTATGTTACAGAAATGCAGACATCGCTGCTTTACGTATCTAGTGG 1320  
QY 1321 GAGAGTGCAGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
Db 1321 GAGAGTGCAGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
QY 1381 TGCCACACTGGGCGAGGAGTGCAGACATCATGTCATCATCTGTCACGTAACTAG 1440  
Db 1381 TGCCACACTGGGCGAGGAGTGCAGACATCATGTCATCATCTGTCACGTAACTAG 1440  
QY 1441 AATACAGAACTTCTCTAATACAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
Db 1441 AATACAGAACTTCTCTAATACAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
QY 1501 TCTACACTGTCAGTACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
Db 1501 TCTACACTGTCAGTACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
QY 1561 TGTCTGAGCGCCCACTTCATGAAATCCCAACAGTGCCTGGTGAAGCGGAAATCTCT 1620  
Db 1561 TGTCTGAGCGCCCACTTCATGAAATCCCAACAGTGCCTGGTGAAGCGGAAATCTCT 1620  
QY 1621 TCACCTTCAGGCGAGAGATGAGTCTGAGTCTAGCTTACGCTTACGAGGCGCGCTCT 1680  
Db 1621 TCACCTTCAGGCGAGAGATGAGTCTGAGTCTAGCTTACGCTTACGAGGCGCGCTCT 1680  
QY 1681 TCGAAGAGAAATGAGAGGCGGACCTCCCGGACTACGATGACCGGATCATTTGCCACCC 1740  
Db 1681 TCGAAGAGAAATGAGAGGCGGACCTCCCGGACTACGATGACCGGATCATTTGCCACCC 1740  
QY 1741 TGAAGCGGTCAGAGCAGCAAGCTGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
Db 1741 TGAAGCGGTCAGAGCAGCAAGCTGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
QY 1801 AGCCCAAAACCAGCTGCGACTGAGTGGGCACTGTGTGGAGAGCGGAGAGACCGCTTGG 1860  
Db 1801 AGCCCAAAACCAGCTGCGACTGAGTGGGCACTGTGTGGAGAGCGGAGAGACCGCTTGG 1860  
QY 1861 AATGTCGTAACCTCTCAGCTTTCGCTCTCATCATTTACCCCGGAGACCGCTTGGTTA 1920  
Db 1861 AATGTCGTAACCTCTCAGCTTTCGCTCTCATCATTTACCCCGGAGACCGCTTGGTTA 1920  
QY 1921 TTTCTCTGAGGTGTGCAACAGGCTCTGCAAGCCCTGGAAGTGGGTGCTCCCGGTGG 1980  
Db 1921 TTTCTCTGAGGTGTGCAACAGGCTCTGCAAGCCCTGGAAGTGGGTGCTCCCGGTGG 1980  
QY 1981 TGTGTGGGGAAGCAGGTCCAGCTTCCCTACAGAGCATGCTGCTGCTGCTGCTGCTGCT 2040  
Db 1981 TGTGTGGGGAAGCAGGTCCAGCTTCCCTACAGAGCATGCTGCTGCTGCTGCTGCTGCT 2040  
QY 2041 TGTGTGTGCCAAAGCTCTGCTGTTACCGAGGTCTATTTCTGCTGCTGCTGCTGCTGCT 2100  
Db 2041 TGTGTGTGCCAAAGCTCTGCTGTTACCGAGGTCTATTTCTGCTGCTGCTGCTGCTGCT 2100

Db 2041 TGTGTGTGCCAAAGCTCTGCTGTTACCGAGGTCTATTTCTGCTGCTGCTGCTGCTGCT 2100  
QY 2101 GTGACCTCTGTGCTATGAGGCGGCAAGGCGCTTTCTGTGGAGACTTACTTCTCAGT 2160  
Db 2101 GTGACCTCTGTGCTATGAGGCGGCAAGGCGCTTTCTGTGGAGACTTACTTCTCAGT 2160  
QY 2161 CTGACAGTATTTTAAATACCGTCTGCTGCTATGATTAAGGACTGCTACAGATCCAGCG 2220  
Db 2161 CTGACAGTATTTTAAATACCGTCTGCTGCTATGATTAAGGACTGCTACAGATCCAGCG 2220  
QY 2221 CTCCATCCGGAAGAGGCGGAGCTGAGATCCCGGCTTACAGGCAAGGCGGCTGGA 2280  
Db 2221 CTCCATCCGGAAGAGGCGGAGCTGAGATCCCGGCTTACAGGCAAGGCGGCTGGA 2280  
QY 2281 CTGACCCCAACATGCTGACAAAGGAGGACCTGAGGCGGCAAGTGCAGAGGAGCGCG 2340  
Db 2281 CTGACCCCAACATGCTGACAAAGGAGGACCTGAGGCGGCAAGTGCAGAGGAGCGCG 2340  
QY 2341 CCTACGCTCTACCCAGATACCTCCGATTTCACTGTGACTGTGACTGCTTTTACCGCA 2400  
Db 2341 CCTACGCTCTACCCAGATACCTCCGATTTCACTGTGACTGTGACTGCTTTTACCGCA 2400  
QY 2401 GCTGGAAGTGTGCTCCAGGCGCTTTCATCTTTCCCGACACTGCTTGAACCTGTGT 2460  
Db 2401 GCTGGAAGTGTGCTCCAGGCGCTTTCATCTTTCCCGACACTGCTTGAACCTGTGT 2460  
QY 2461 TGCCCTCAGAGGCAAAATCTTGGGCTCAGGAGCTGCTTGGGCTATTTGGTGGTAG 2520  
Db 2461 TGCCCTCAGAGGCAAAATCTTGGGCTCAGGAGCTGCTTGGGCTATTTGGTGGTAG 2520  
QY 2521 CTGGGGGTTCTGGCAAGGAATTCAGGCGAGGCTTGGAGGCAATTTCCCGAGAGCA 2580  
Db 2521 CTGGGGGTTCTGGCAAGGAATTCAGGCGAGGCTTGGAGGCAATTTCCCGAGAGCA 2580  
QY 2581 TCACGCTGTGATGTTGATTAATGAGCGGAGAGAGTGTGTAATCTTTAGAGAGAG 2640  
Db 2581 TCACGCTGTGATGTTGATTAATGAGCGGAGAGAGTGTGTAATCTTTAGAGAGAG 2640  
QY 2641 TGAATGCGCTCCCTTACCTGAACAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700  
Db 2641 TGAATGCGCTCCCTTACCTGAACAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700  
QY 2701 CAGAGGACCTTCTGCGGCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760  
Db 2701 CAGAGGACCTTCTGCGGCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760  
QY 2761 ACAGTTTGAACAACCGATCTTACCTGGAATGGAATGAGACAGAGGCAATCCGCTCA 2820  
Db 2761 ACAGTTTGAACAACCGATCTTACCTGGAATGGAATGAGACAGAGGCAATCCGCTCA 2820  
QY 2821 TTGATGACGATGCTACCTCCGCAATGCAAGAAATCATGTTTGGTTCGGGTGTGAGAG 2880  
Db 2821 TTGATGACGATGCTACCTCCGCAATGCAAGAAATCATGTTTGGTTCGGGTGTGAGAG 2880  
QY 2881 AAGCTCGGAGCGGATGCTGCGGCTTCCCTGCGCTTACAGGCAATGGAGCATTCGCCAT 2940  
Db 2881 AAGCTCGGAGCGGATGCTGCGGCTTCCCTGCGCTTACAGGCAATGGAGCATTCGCCAT 2940  
QY 2941 AGTCTGCTCTACAACTCACTACTCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 3000  
Db 2941 AGTCTGCTCTACAACTCACTACTCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 3000  
QY 3001 CCTTCTTCAACAGTATTAATGCTACCTGATTTCTATGATGATCCCGGAGGCAATCCGG 3060  
Db 3001 CCTTCTTCAACAGTATTAATGCTACCTGATTTCTATGATGATCCCGGAGGCAATCCGG 3060  
QY 3061 ACATGCTGATGATTAATTAATCACTGTGAGAGCATTTGCTGCTGCTGCTGCTGCTGCT 3120  
Db 3061 ACATGCTGATGATTAATTAATCACTGTGAGAGCATTTGCTGCTGCTGCTGCTGCTGCT 3120  
QY 3121 TCACTGGAAGCCCGCATCAAGGTGACCTCAGGCTGGAATTCGATTCGCTGCTGCTGCTGCT 3180  
Db 3121 TCACTGGAAGCCCGCATCAAGGTGACCTCAGGCTGGAATTCGATTCGCTGCTGCTGCTGCT 3180

(E-mail: cdhain@fokazusa.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)  
 2 (sites)  
 Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.  
 Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro  
 Coder for large proteins in vitro  
 DNA Res. 5 (1), 31-39 (1998)  
 989050545

JOURNAL  
 MEDLINE  
 FEATURES  
 source

TITLE

REFERENCE  
 AUTHORS

1. 6189  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HG1237"  
 /sex="male"  
 /tissue\_type="brain"  
 /clone\_id="pbluescriptII SK plus"  
 616..3376  
 /gene="KIAA0519"  
 616..3375  
 /gene="KIAA0519"  
 /product="1"  
 /codon\_start=1  
 /protein\_id="BAA25445.1"  
 /db\_xref="GI:30433562"  
 /translation="MTGYTMLRNGGANGGONCMLRWSNRIRLWLSFTLVIVLFFPL  
 LNHVYLLTLDDEAGKRIFGPRNGNELCEKHYLDICRIESESSELDLEAKRO  
 LNEIKYLLTDEKTEACKSTENAKODLLDKNYSISOTESHYKELMAONPKLSLRLRL  
 PEKDDAGLPKPRATRCGRHLNCFDYSRCLVTSGFVYVYSDQFVFGSYLDPLVQA  
 OATRAANVYVENADTACLVYLVEEMQEVVLRARLEKQLSLPHRTPGHNHVI  
 NLSRKDTONLLYNWSTGRANVAQSTFYVQVRFEDLVSPVLHAMESEPMELP  
 VPKRFLYFTFOGKPIESLRSSLOJAREPDLVSPVLHAMESEPMELP  
 OYLVKFTCKNPKPSLPTFMAALGGERDEBELKSTFALITTPGDRPLVSSCATR  
 LBEALVGAQVAVLGEVOLPYOMDLONNEALVYKRYATVEHFLSLDSOLLMA  
 ROAGRLMETVPTADSIENYVLAIRTRIDIPAPVINEEAAEIPKRSKAACTDNP  
 MANDDLIDIGPVEETPPYASPRYLKNETLVITVDEYRSWNAKGFPHLEPFPDPLP  
 SAKPLKSGSTGRPLTGGGAGSGGKEFOALGDNPRQETVMLTYEEVEYLNMSLE  
 LNGDLYLVVYVWNSPKLPSLEDLMPDIGVIMVYRTEKSLNNRFLPMNETEAL  
 LSIIDDAHLRHDEIMFGRVYREARDRVGFPGRYHAMVDIHOSSMLVNSYSCLSNV  
 LTGAFLKFRYYAVLYSYVPOAIRMDEVEYINCEDIAMNLFVSHITTKRPIKTSMT  
 CFKPKCPALSHSDSHFHERHKCINFPKVVGYMPLYTORVDVSVLEKTRLPDKTK

BASE COUNT 1279 a 1687 c 1706 g 1517 t

ORIGIN

Query Match 99.9%; Score 6163.8; DB 9; Length 6189;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 6165; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGGCTCCTAGCTGGAAGCCGGAGCAAGCCCTGGAGGTTACCTCTTCAAGAAGT 60  
 |||||  
 Db 23 GGGGGGCTCCTAGCTGGAAGCCGGAGCAAGCCCTGGAGGTTACCTCTTCAAGAAGT 82

QY 61 CGTGTGCTAGAGGTGAAGCTACACAACACTCAGAGAGGAAGAGGCTCTGAAACATATGC 120  
 |||||  
 Db 83 CGTGTGCTAGAGGTGAAGCTACACAACACTCAGAGAGGAAGAGGCTCTGAAACATATGC 142

QY 121 CTGATTTGTGGCAAGGCATCATAGAAGCTGCGCATTTATTTCTGTTCTTAACCTATTACT 180  
 |||||  
 Db 143 CTGATTTGTGGCAAGGCATCATAGAAGCTGCGCATTTATTTCTGTTCTTAACCTATTACT 202

QY 181 GTATAAGCTGAATAAGACATATGCATATTTGTTGGTCAGCAAAACCAAGAAACAAGAGC 240  
 |||||  
 Db 203 GTATAAGCTGAATAAGACATATGCATATTTGTTGGTCAGCAAAACCAAGAAACAAGAGC 262

QY 241 TATGGCATTTGAAAAAATCTGTCTGATTTCCAGGGTGTGTTTCTGGGTTTTCATCATCAGG 300  
 |||||  
 Db 263 TATGGCATTTGAAAAAATCTGTCTGATTTCCAGGGTGTGTTTCTGGGTTTTCATCATCAGG 322

QY 301 TACTCTCCCTTTCATCTCAGCAAGAAATGTGGCACTTTTATCTGTTGATTAAGATTAA 360

Oy	3181	CTGAGGCCCTGTCTCATGATGACTCCCACTTCCACAGCGGGCAAGTGCATCAACTCTT	3240
Dp	3181	CTCAGGGCCCTGTCTCATGATGACTCCCACTTCCACAGCGGGCAAGTGCATCAACTCTT	3240
Oy	3241	TGCGAAGGTGTACGGGTATCATGCCCCCTCTGTACACGAGTTCCAGGGGTGATTTCTGTC	3300
Dp	3241	TGCGAAGGTGTACGGGTATCATGCCCCCTCTGTACACGAGTTCCAGGGGTGATTTCTGTC	3300
Oy	3301	TCTTCAAGACACGGCTTGGCCCATGACAGCAACAACTGCTTCAAGTTCATCTAGGGGACGC	3360
Dp	3301	TCTTCAAGACACGGCTTGGCCCATGACAGCAACAACTGCTTCAAGTTCATCTAGGGGACGC	3360
Oy	3361	GCACGGGTCTGGGGAAGAGGATGTAGACAGAGGGGAAGATGGCTCCCAAGTTCCTTAGGCA	3420
Dp	3361	GCACGGGTCTGGGGAAGAGGATGTAGACAGAGGGGAAGATGGCTCCCAAGTTCCTTAGGCA	3420
Oy	3421	TTTGAGGACCTTGGGCACTCTCTGGTGGGTGGCCCAAGGCTCTGCTGGAAAGGGGCG	3480
Dp	3421	TTTGAGGACCTTGGGCACTCTCTGGTGGGTGGCCCAAGGCTCTGCTGGAAAGGGGCG	3480
Oy	3481	CAGAGAGGTGTGAAGAAACCCGTGCTTATCTTAAATCAGCCACACTGGGCTTGAG	3540
Dp	3481	CAGAGAGGTGTGAAGAAACCCGTGCTTATCTTAAATCAGCCACACTGGGCTTGAG	3540
Oy	3541	CCCTGGGCGGAGTCCCGGGGGTTCCCAACAGGGGACATGATAGCTTACACTGAGG	3600
Dp	3541	CCCTGGGCGGAGTCCCGGGGGTTCCCAACAGGGGACATGATAGCTTACACTGAGG	3600
Oy	3601	ACTGTGGGACCTGTGAGAGTCACTACACACCGTGTACACCCAGACAGCTGGTCTGTG	3660
Dp	3601	ACTGTGGGACCTGTGAGAGTCACTACACACCGTGTACACCCAGACAGCTGGTCTGTG	3660
Oy	3661	GTTTTCACATTCATAACACTATATGATATTTTAAAGAGAAAGTTTCAGATTTGCC	3720
Dp	3661	GTTTTCACATTCATAACACTATATGATATTTTAAAGAGAAAGTTTCAGATTTGCC	3720
Oy	3721	ATTCAGGCTTATTTATATATATGTGTATATTAATATACATCACAACACTTGCATACA	3780
Dp	3721	ATTCAGGCTTATTTATATATATGTGTATATTAATATACATCACAACACTTGCATACA	3780
Oy	3781	TATATATTTTGGCTGGGGAGGTGTAGATTTTGGCTTAAAGGAGGGACCGGCCAGGC	3840
Dp	3781	TATATATTTTGGCTGGGGAGGTGTAGATTTTGGCTTAAAGGAGGGACCGGCCAGGC	3840
Oy	3841	TCTTTGTCTGTATATCTGTGGCGAGATGGGTCTGTGCCCTTGTCTACTGCTTATCTTA	3900
Dp	3841	TCTTTGTCTGTATATCTGTGGCGAGATGGGTCTGTGCCCTTGTCTACTGCTTATCTTA	3900
Oy	3901	AAGATCATCTCCATCTCTCCCAAGGCGCATCTGTGTGACAGCAACAGAAAGGATGAAC	3960
Dp	3901	AAGATCATCTCCATCTCTCCCAAGGCGCATCTGTGTGACAGCAACAGAAAGGATGAAC	3960
Oy	3961	TGGGCCCTTTGGGGCCCTGTGACAAGGTCTCTTCCATACCTTTCTGTTGCCAGTAGCAA	4020
Dp	3961	TGGGCCCTTTGGGGCCCTGTGACAAGGTCTCTTCCATACCTTTCTGTTGCCAGTAGCAA	4020
Oy	4021	CTGTAACTACATCTCTTCCAGTGAATCCCTGGGAGCGCTCAACCTGGTGGGCTGT	4080
Dp	4021	CTGTAACTACATCTCTTCCAGTGAATCCCTGGGAGCGCTCAACCTGGTGGGCTGT	4080
Oy	4081	TACACTTCGCTGTGGGGCCAGGCAATTTTGGAGATTTATCTTAAAGCCAGGGTTGCC	4140
Dp	4081	TACACTTCGCTGTGGGGCCAGGCAATTTTGGAGATTTATCTTAAAGCCAGGGTTGCC	4140
Oy	4141	TCCGTACTTATCCCTGTCTCCCAATTTCTCTTGTTTGAGAGAAATGAGGAAGCAAG	4200
Dp	4141	TCCGTACTTATCCCTGTCTCCCAATTTCTCTTGTTTGAGAGAAATGAGGAAGCAAG	4200
Oy	4201	AGTGAAGAAATATGAGGGGTGGAAGAGCCACATCCAGATGGGTCTTTTATCTGTCTTT	4260
Dp	4201	AGTGAAGAAATATGAGGGGTGGAAGAGCCACATCCAGATGGGTCTTTTATCTGTCTTT	4260

QY	4261	CTGTGAAACACACGTGCTGTGGGCCCTCAGAGCCTTTCTGAAGTGTCTTCTTGGAATGG	43220
Db	4261	CTGTGAAACACACGTGCTGTGGGCCCTCAGAGCCTTTCTGAAGTGTCTTCTTGGAATGG	43200
QY	4321	ACAGAGATGTGACGACGCTGGACATCTCTGTGGTCTGAAGTGATGGTTGACAGTCACTCC	43800
Db	4321	ACAGAGATGTGACGACGCTGGACATCTCTGTGGTCTGAAGTGATGGTTGACAGTCACTCC	43800
QY	4381	TCTCCCTAGTGTAGAGCAAGCAGATGTCTTCAGAGAACCCACCGGCTGGCCGGGAAGT	44400
Db	4381	TCTCCCTAGTGTAGAGCAAGCAGATGTCTTCAGAGAACCCACCGGCTGGCCGGGAAGT	44400
QY	4441	TTTACAGCAAGGCCCTGTGCTGGGATTAATCTCTTGTAATTAACCTTCCCCCGCT	45000
Db	4441	TTTACAGCAAGGCCCTGTGCTGGGATTAATCTCTTGTAATTAACCTTCCCCCGCT	45000
QY	4501	CTGTGTGAGGCCCATCTCGTGTATCTGTGTGGTGGTGGCCCTAATGTACGTTGGCT	45600
Db	4501	CTGTGTGAGGCCCATCTCGTGTATCTGTGTGGTGGTGGCCCTAATGTACGTTGGCT	45600
QY	4561	GTAGAGACCCCCGAGAGTTTGGTATGTGCTAGAACAAATGGGAGGCTGTGATTTGCTGTGA	46200
Db	4561	GTAGAGACCCCCGAGAGTTTGGTATGTGCTAGAACAAATGGGAGGCTGTGATTTGCTGTGA	46200
QY	4621	AGCTCACATCCAGCCTTGGATCTAACGGGCATTCACAAACCGAGTTTACCATTTCCTACT	46800
Db	4621	AGCTCACATCCAGCCTTGGATCTAACGGGCATTCACAAACCGAGTTTACCATTTCCTACT	46800
QY	4681	CCCTCCTTAAGATTTCTGTGCCCTGGGGGTGAATCTGAATAATAGCTAATTTTGGGTACAG	47400
Db	4681	CCCTCCTTAAGATTTCTGTGCCCTGGGGGTGAATCTGAATAATAGCTAATTTTGGGTACAG	47400
QY	4741	GTGGCAGTAGGGGAACCTAGAGAGGGTGTGAGTGGCATTTGTACAGGATTTAGCCCATGAC	48000
Db	4741	GTGGCAGTAGGGGAACCTAGAGAGGGTGTGAGTGGCATTTGTACAGGATTTAGCCCATGAC	48000
QY	4801	GTGTCTTCTGAACCTTACTTTCTGTGAAGTGTGAATTAATCTGTGAAGTTTCTAGAACATG	48600
Db	4801	GTGTCTTCTGAACCTTACTTTCTGTGAAGTGTGAATTAATCTGTGAAGTTTCTAGAACATG	48600
QY	4861	AACAAACCTAGAGTTTGTCTCGGTATGTGACATGTCATTAAGCAGTTCCGCTTCCCTA	49200
Db	4861	AACAAACCTAGAGTTTGTCTCGGTATGTGACATGTCATTAAGCAGTTTCCGCTTCCCTA	49200
QY	4921	GACCTTGGCATCTGTGCTTCTAATTTCTTGGAATAGTTTCTCTGTACCTGCTGTAC	49800
Db	4921	GACCTTGGCATCTGTGCTTCTAATTTCTTGGAATAGTTTCTCTGTACCTGCTGTAC	49800
QY	4981	ACGTGGGCTCTTCCAAAGTACTGTTTGAAGCTGGGCTCTTTGTGTAGTCTCCACCCAC	50400
Db	4981	ACGTGGGCTCTTCCAAAGTACTGTTTGAAGCTGGGCTCTTTGTGTAGTCTCCACCCAC	50400
QY	5041	CTGTAGGGCTAGCTCGGCTTAAGGGAATCTCCCAATTTGGCAAAACCGGACCGGCGCGC	51000
Db	5041	CTGTAGGGCTAGCTCGGCTTAAGGGAATCTCCCAATTTGGCAAAACCGGACCGGCGCGC	51000
QY	5101	CCAGAGCATGTCTTCCAAAGGTTCCCGGCCCCCAACCCAGCATCAGCCTGTAGTCTCCC	51600
Db	5101	CCAGAGCATGTCTTCCAAAGGTTCCCGGCCCCCAACCCAGCATCAGCCTGTAGTCTCCC	51600
QY	5161	TGCTGAGGCACTGTGTATGTTCGCCAGCATGTGGGGGTGAGAGCGCCCTTCCACAACTT	52200
Db	5161	TGCTGAGGCACTGTGTATGTTCGCCAGCATGTGGGGGTGAGAGCGCCCTTCCACAACTT	52200
QY	5221	TCTATGTTCCCTTACCTGACGATCTGTGCTTATCTTTTGTGACATGACCTTCTTCCT	52800
Db	5221	TCTATGTTCCCTTACCTGACGATCTGTGCTTATCTTTTGTGACATGACCTTCTTCCT	52800
QY	5281	CGGGAGCAAAAGATGTGCTGTGTGGACCTAATTTGTGGTCAATTTCAATCTTGTTTC	53400
Db	5281	CGGGAGCAAAAGATGTGCTGTGTGGACCTAATTTGTGGTCAATTTCAATCTTGTTTC	53400
QY	5341	TTTTTAATGTGGAATCTCACATACTGATTCAGTGTGGACTCGGTGAGCCGGGCCGTGTG	54000

Db 323 TACCTCTCCCTTTTCAATCTCAGCAAGATGTGGCACCTTTTATCGTTTGATTAAGATTAA 382  
Qy 361 GCACATGTTCTTGTGTCAACAGCCAGAACTTAAATCTGCTGGAAATAGGTCAGAGACA 420  
Db 383 GGCATGTTCTTGTGTCAACAGCCAGAACTTAAATCTGCTGGAAATAGGTCAGAGACA 442  
Qy 421 TTTTACGCTGACGTGAGAAATGAATGTCTATTTTATTTGGTGGCTTGTCTGGGAGC 480  
Db 443 TTTTACGCTGACGTGAGAAATGAATGTCTATTTTATTTGGTGGCTTGTCTGGGAGC 502  
Qy 481 ACCTAACCTCTTGGAAACGTGTCTGAAACAGATCGTTTGTGGAAATAGGTCAGAGACC 540  
Db 503 ACCTAACCTCTTGGAAACGTGTCTGAAACAGATCGTTTGTGGAAATAGGTCAGAGACC 562  
Qy 541 ATGATTATGGCAGATGACCCGACGTATCTGGGGCAGGCTCAGAGACTCATGACAG 600  
Db 563 ATGATTATGGCAGATGACCCGACGTATCTGGGGCAGGCTCAGAGACTCATGACAG 622  
Qy 601 GCTATACCATGCTGGCGAATGGGGCGGGGGAACGAGAGCTCAGACCTGCTGCGCT 660  
Db 623 GCTATACCATGCTGGCGAATGGGGCGGGGGAACGAGAGCTCAGACCTGCTGCGCT 682  
Qy 661 GGTCCAACCGCATCGGCTCAGCTGGGTGCTCAGGCTTTTGTATCTGCTGCTTCT 720  
Db 683 GGTCCAACCGCATCGGCTCAGCTGGGTGCTCAGGCTTTTGTATCTGCTGCTTCT 742  
Qy 721 TCCGCTCATCGCCACCTATTACCTCAGCACTCTGGATGAGGCTGATGAGGCAAGC 780  
Db 743 TCCGCTCATCGCCACCTATTACCTCAGCACTCTGGATGAGGCTGATGAGGCAAGC 802  
Qy 781 GGAATTTTGGTCCCCGGGTGGGGAAGAGCTGTGCGAGGTGAAGACAGCTGCTGATCTGT 840  
Db 803 GGAATTTTGGTCCCCGGGTGGGGAAGAGCTGTGCGAGGTGAAGACAGCTGCTGATCTGT 862  
Qy 841 GCGCATCCGCGAGTGGGTGAGTGAAGAGCTCTGCGAGCTGGAGGCAAGCCGCAAGAGC 900  
Db 863 GCGCATCCGCGAGTGGGTGAGTGAAGAGCTCTGCGAGCTGGAGGCAAGCCGCAAGAGC 922  
Qy 901 TGAACAGCGAGATCGCCAACTGAATCTGAAGATCGAAGCCTGTAAAGAGCATTTGAGA 960  
Db 923 TGAACAGCGAGATCGCCAACTGAATCTGAAGATCGAAGCCTGTAAAGAGCATTTGAGA 982  
Qy 961 ACCCGAAGCAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGAGCATTTCTACA 1020  
Db 983 ACCCGAAGCAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGAGCATTTCTACA 1042  
Qy 1021 AGGAGCTCATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGACCTGCTCCAGAGA 1080  
Db 1043 AGGAGCTCATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGACCTGCTCCAGAGA 1102  
Qy 1081 AGGAGCTCATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGACCTGCTCCAGAGCTGT 1140  
Db 1103 AGGAGCTCATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGACCTGCTCCAGAGCTGT 1162  
Qy 1141 TTTATTTATTTCTGCTTGGCTCTCACTCTGCTGCTTCCGGCTTCAAGCTTATGACAGTGACC 1200  
Db 1163 TTTATTTATTTCTGCTTGGCTCTCACTCTGCTGCTTCCGGCTTCAAGCTTATGACAGTGACC 1222  
Qy 1201 AGTTTGCTTTGGCAGCTACCTGATCCCTTGGTCAAGAGGCTTTTTCAGGCGACAGCAC 1260  
Db 1223 AGTTTGCTTTGGCAGCTACCTGATCCCTTGGTCAAGAGGCTTTTTCAGGCGACAGCAC 1282  
Qy 1261 GAGCTAACGTTTATGTTTACAGAAATGTCAGACATCGCTGCTTTTACGTTATCTAGTGG 1320  
Db 1283 GAGCTAACGTTTATGTTTACAGAAATGTCAGACATCGCTGCTTTTACGTTATCTAGTGG 1342  
Qy 1321 GAGAGATGACAGACCCGCTGCTGCGGCTCTGCTGAGCTGAGAAACAGATTTATTTCC 1380  
Db 1343 GAGAGATGACAGACCCGCTGCTGCGGCTCTGCTGAGCTGAGAAACAGATTTATTTCC 1402  
Qy 1381 TGGCACACCTGGCGAGAGGATGAGACACACCATGTCATCATCATCTGTCACGTAAGTCAG 1440  
Db 1403 TGGCACACCTGGCGAGAGGATGAGACACACCATGTCATCATCATCTGTCACGTAAGTCAG 1462

Qy 1441 ATACACAGAACCTTCTCTATTAACGTCACTAGGCGCGTGCATGTTGGCCAGTCCACCT 1500  
Db 1463 ATACACAGAACCTTCTCTATTAACGTCACTAGGCGCGTGCATGTTGGCCAGTCCACCT 1522  
Qy 1501 TCTACACTGTCCAGATTCACACCTGGCTTTGACTTGGTCTATTCACCGCTGTGTCATGCCA 1560  
Db 1523 TCTACACTGTCCAGATTCACACCTGGCTTTGACTTGGTCTATTCACCGCTGTGTCATGCCA 1582  
Qy 1561 TGTGTAGGCCCAACTTCATGAAATCCACACAGATGGCGGGGAAGCGAAATATCTCT 1620  
Db 1583 TGTGTAGGCCCAACTTCATGAAATCCACACAGATGGCGGGGAAGCGAAATATCTCT 1642  
Qy 1621 TCACCTTCCAGGGCGAGAAAGATTGACTCTGTAGAGTCTAGCTTTCAGAGGCCCGCTCT 1680  
Db 1643 TCACCTTCCAGGGCGAGAAAGATTGACTCTGTAGAGTCTAGCTTTCAGAGGCCCGCTCT 1702  
Qy 1681 TCGAAGAGAAATGAGAGGCGACCTCCCGCGGACACTACGATGACCGGATCATTTGCCACCC 1740  
Db 1703 TCGAAGAGAAATGAGAGGCGACCTCCCGCGGACACTACGATGACCGGATCATTTGCCACCC 1762  
Qy 1741 TGAAGCGGTGACAGACAGCAAGCTGATCAGGTCTGTGTAATTCACCTGCAGAAAACC 1800  
Db 1763 TGAAGCGGTGACAGACAGCAAGCTGATCAGGTCTGTGTAATTCACCTGCAGAAAACC 1822  
Qy 1801 AGCCCAACCCAGACCTGCCGACTGAGTGGGACACTGTGTGAGAGCGGGAGAGACCGCTTGG 1860  
Db 1823 AGCCCAACCCAGACCTGCCGACTGAGTGGGACACTGTGTGAGAGCGGGAGAGACCGCTTGG 1882  
Qy 1861 AATTGCTGAAGCTCTCCACCTTCGCGCTCATCATATTCACCCCGGGGACCCCTCGCTTGGTTA 1920  
Db 1883 AATTGCTGAAGCTCTCCACCTTCGCGCTCATCATATTCACCCCGGGGACCCCTCGCTTGGTTA 1942  
Qy 1921 TTTCTCTGTGGTGTGCAACACAGGCTCTTGGAAAGCCCTGGAAGTGGTGGCTCCGCTGG 1980  
Db 1943 TTTCTCTGTGGTGTGCAACACAGGCTCTTGGAAAGCCCTGGAAGTGGTGGCTCCGCTGG 2002  
Qy 1981 TGTGGGGAGCAGGTTCACACTTCCCTACCAAGACATGCTGCAGTGTGAACAGAGCGGCC 2040  
Db 2003 TGTGGGGAGCAGGTTCACACTTCCCTACCAAGACATGCTGCAGTGTGAACAGAGCGGCC 2062  
Qy 2041 TGTGGTGGCAAGGCTCGGTGACGAGGTTTCCGTGCTCAGAAAGCTCTCCGATTA 2100  
Db 2063 TGTGGTGGCAAGGCTCGGTGACGAGGTTTCCGTGCTCAGAAAGCTCTCCGATTA 2122  
Qy 2101 GTGACCTCTGTGCTATGAGGCGGCAAGCGCGCTTCTCTGSGAGACTTACTTCTCCACTG 2160  
Db 2123 GTGACCTCTGTGCTATGAGGCGGCAAGCGCGCTTCTCTGSGAGACTTACTTCTCCACTG 2182  
Qy 2161 CTGACAGTATTTTAAATPACCGTGTGCTGATGATGAGACTGCGATCCAGATCCAGCCG 2220  
Db 2183 CTGACAGTATTTTAAATPACCGTGTGCTGATGATGAGACTGCGATCCAGATCCAGCCG 2242  
Qy 2221 CTGCCATCCGGGGAAGAGGGGAGCTGAGATCCGCCACCGTTTACGCGAAGCGCGCTGGA 2280  
Db 2243 CTGCCATCCGGGGAAGAGGGGAGCTGAGATCCGCCACCGTTTACGCGAAGCGCGCTGGA 2302  
Qy 2281 CTGACCCCAACATGAGTGAACAGGGGAGCTGAGACTGGGGCCAGTGGAGACGAGACCCG 2340  
Db 2303 CTGACCCCAACATGAGTGAACAGGGGAGCTGAGACTGGGGCCAGTGGAGACGAGACCCG 2362  
Qy 2341 CTTAGCCTTCACCCAGATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2400  
Db 2363 CTTAGCCTTCACCCAGATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2422  
Qy 2401 GCTGGAATGTGTCTCAGAGGCTTTTCCATCTTTTCCCCCAACTCCCTTTGACACCTGTGT 2460  
Db 2423 GCTGGAATGTGTCTCAGAGGCTTTTCCATCTTTTCCCCCAACTCCCTTTGACACCTGTGT 2482  
Qy 2461 TGGCCTCAGAGGCAATTTCTTGGGCTCAGGAGCTGGCTTTTGGGCTATTTGGGTGGAG 2520  
Db 2483 TGGCCTCAGAGGCAATTTCTTGGGCTCAGGAGCTGGCTTTTGGGCTATTTGGGTGGAG 2542

QY 2521 CTGGGGGCTTCGCAAGAAATTCAGGACGCTTGAGGCAATGTTCCCGAGAGCACT 2580  
|||||  
Db 2543 CTGGGGGCTTCGCAAGAAATTCAGGACGCTTGAGGCAAGTTCCCGAGAGCACT 2602  
QY 2581 TCACGGTGTGATGTGACTTATGAGCGGAGAGAGTCTTATGAATCTTTTGAGAGGC 2640  
|||||  
Db 2603 TCACGGTGTGATGTGACTTATGAGCGGAGAGAGTCTTATGAATCTTTTGAGAGGC 2662  
QY 2641 TGAATGGCCCTCCCTACCCGAAACAGAGTCTGTGTGGAAATTTCTCCAGCTGCCAT 2700  
|||||  
Db 2663 TGAATGGCCCTCCCTACCCGAAACAGAGTCTGTGTGGAAATTTCTCCAGCTGCCAT 2722  
QY 2701 CAGAGACCTTCTGTGGCTGACATTTGGGCTTCCATCATGAGTGGCCGCTACTGAGAGA 2760  
|||||  
Db 2723 CAGAGACCTTCTGTGGCTGACATTTGGGCTTCCATCATGAGTGGCCGCTACTGAGAGA 2782  
QY 2761 ACAGTTGAACAACCGATTTCTTACCTGTGAATGAAATTTGAGACAGAGGCACTCTGTCA 2820  
|||||  
Db 2783 ACAGTTGAACAACCGATTTCTTACCTGTGAATGAAATTTGAGACAGAGGCACTCTGTCA 2842  
QY 2821 TTGATGACGATGCTCACTCCGCGCAATGAGAAATCATGTTTGGGTTCCGGGTGGAGAG 2880  
|||||  
Db 2843 TTGATGACGATGCTCACTCCGCGCAATGAGAAATCATGTTTGGGTTCCGGGTGGAGAG 2902  
QY 2881 AAGCTCGGAGACCGCATCTGTGGCTTCCGCGCTTACACGCAATGGAGCATCCCCCATC 2940  
|||||  
Db 2903 AAGCTCGGAGACCGCATCTGTGGCTTCCGCGCTTACACGCAATGGAGCATCCCCCATC 2962  
QY 2941 AGTCTGGCTCTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGAGTGTG 3000  
|||||  
Db 2963 AGTCTGGCTCTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGAGTGTG 3022  
QY 3001 CCTCTTTCACAAGATATATGCTACCTGATTTCTTATGATGATGCCAGCCATCCGG 3060  
|||||  
Db 3023 CCTCTTTCACAAGATATATGCTACCTGATTTCTTATGATGATGCCAGCCATCCGG 3082  
QY 3061 ACATGTTGATGATATACATCACTGTGAGGACATTTGCCATGAATCTCTTGTCTCCACA 3120  
|||||  
Db 3083 ACATGTTGATGATATACATCACTGTGAGGACATTTGCCATGAATCTCTTGTCTCCACA 3142  
QY 3121 TCACCTCGAAGCCCGCATAGGTGACCTCAGCGTGTGACATTCGATGCCAGATGCC 3180  
|||||  
Db 3143 TCACCTCGAAGCCCGCATAGGTGACCTCAGCGTGTGACATTCGATGCCAGATGCC 3202  
QY 3181 CTCAGGCCCTGTCTATGATGATCCCATCTCCACGAGGGGCAACATGATCAACTTCT 3240  
|||||  
Db 3203 CTCAGGCCCTGTCTATGATGATCCCATCTCCACGAGGGGCAACATGATCAACTTCT 3262  
QY 3241 TCGTGAAGGTGTACGGCTACATGCCCTCTCTGTACACGCAAGTTCAGGTTGATTTGTGC 3300  
|||||  
Db 3263 TCGTGAAGGTGTACGGCTACATGCCCTCTCTGTACACGCAAGTTCAGGTTGATTTGTGC 3322  
QY 3301 TCTTCAAGACAGCGCTGCCCATGACAAAGCAAGTGTTCAAAGTTCATCTAGGGGACAG 3360  
|||||  
Db 3323 TCTTCAAGACAGCGCTGCCCATGACAAAGCAAGTGTTCAAAGTTCATCTAGGGGACAG 3382  
QY 3361 GCAAGGTTGAGGAAAGAGATGAGCAGAGGAGAAAGTGGCTCCCAAGTTCTTAGGCA 3420  
|||||  
Db 3383 GCAAGGTTGAGGAAAGAGATGAGCAGAGGAGAAAGTGGCTCCCAAGTTCTTAGGCA 3442  
QY 3421 TTGCAAGACCTTTGGGACATCTGTGTGGTGGTGGCCACAGGCTGTCTGGAAGGGGAG 3480  
|||||  
Db 3443 TTGCAAGACCTTTGGGACATCTGTGTGGTGGTGGCCACAGGCTGTCTGGAAGGGGAG 3502  
QY 3481 CAGGAGAGTGAAGAAACCGCTCCCTTATCTTGAAGTCAAGCCATGGGCTGTGAG 3540  
|||||  
Db 3503 CAGGAGAGTGAAGAAACCGCTCCCTTATCTTGAAGTCAAGCCATGGGCTGTGAG 3562  
QY 3541 CCTGGGGGAGATCCCGGGGTTCCCAACAGGSCATGACTGATTAACCTTAGAGG 3600  
|||||  
Db 3563 CCTGGGGGAGATCCCGGGGTTCCCAACAGGSCATGACTGATTAACCTTAGAGG 3622  
QY 3601 ACTGTGGGACTGTGCAAGTCACTACACCGTTCTGTACGCCAGAGACAGTGGTCTGTG 3660  
|||||

Db 3623 ACTGTGGGACTGTGCAAGTCACTACACCGTTCTGTACGCCAGAGACAGCTGGTCTGTG 3682  
QY 3661 GTTTTACATTCAATAACAATATATGATTTATTTAAAGAGAAATTTCAATTTGGC 3720  
|||||  
Db 3683 GTTTTACATTCAATAACAATATATGATTTATTTAAAGAGAAATTTCAATTTGGC 3742  
QY 3721 ATTCAAGGCTTATTTATATATATGTGTGTATATATAATACATGACACACTTGTAC 3780  
|||||  
Db 3743 ATTCAAGGCTTATTTATATATATGTGTGTATATATAATACATGACACACTTGTAC 3802  
QY 3781 TATATATTTTGGCTGGGAGAGTGTGATTTGGCTTTCTAAGGAGGAGACCGCGAGC 3840  
|||||  
Db 3803 TATATATTTTGGCTGGGAGAGTGTGATTTGGCTTTCTAAGGAGGAGACCGCGAGC 3862  
QY 3841 TCCTTTGTCTGTATTTGTGGAGAGATGGGCTGTGGCTTTGTACATGCTTATCTTTA 3900  
|||||  
Db 3863 TCCTTTGTCTGTATTTGTGGAGAGATGGGCTGTGGCTTTGTACATGCTTATCTTTA 3922  
QY 3901 AAGATCATCTCCCATCCATCCCGAGCCCATCTGTGTGACAGCAACGAGAAAGGATGAAT 3960  
|||||  
Db 3923 AAGATCATCTCCCATCCATCCCGAGCCCATCTGTGTGACAGCAACGAGAAAGGATGAAT 3982  
QY 3961 TGGCCCTTGTGGGGGCTGAGCAAGGTCCTTCTTACCTTTCTGTGGCACTAGCAA 4020  
|||||  
Db 3983 TGGCCCTTGTGGGGGCTGAGCAAGGTCCTTCTTACCTTTCTGTGGCACTAGCAA 4042  
QY 4021 CCTGTACTCAATTTCTTCTCCAGTGAATCCCTGGAGGCGCTGTACCCCTGGTGGGCTGT 4080  
|||||  
Db 4043 CCTGTACTCAATTTCTTCTCCAGTGAATCCCTGGAGGCGCTGTACCCCTGGTGGGCTGT 4102  
QY 4081 TCAGTTCCTGCTGTGGGGGCAAGCATTTTGTAGAGATTTATCTTTAGGCGAGCTTGGC 4140  
|||||  
Db 4103 TCAGTTCCTGCTGTGGGGGCAAGCATTTTGTAGAGATTTATCTTTAGGCGAGCTTGGC 4162  
QY 4141 TCGTACTTATCCCTGTCTCCATTTCTCTTCTTGTGAGAGATTTAGAGAAAGCAAG 4200  
|||||  
Db 4163 TCGTACTTATCCCTGTCTCCATTTCTCTTCTTGTGAGAGATTTAGAGAAAGCAAG 4222  
QY 4201 AGTGAAGAAATATAGGGGCTGGAAGCGCACTCCAGATGGCTTCTTATCCCTGTCT 4260  
|||||  
Db 4223 AGTGAAGAAATATAGGGGCTGGAAGCGCACTCCAGATGGCTTCTTATCCCTGTCT 4282  
QY 4261 CTGTGAAACACAGTGTGTGGGCTCAGGCGCTTCTGAAAGTCTTCTTGTGATGG 4320  
|||||  
Db 4283 CTGTGAAACACAGTGTGTGGGCTCAGGCGCTTCTGAAAGTCTTCTTGTGATGG 4342  
QY 4321 ACAGGAGATCAGCAGGTGCATCTGTGTGTGAAAGTGTGTGCAAGTCAAGCTCC 4380  
|||||  
Db 4343 ACAGGAGATCAGCAGGTGCATCTGTGTGTGAAAGTGTGTGCAAGTCAAGCTCC 4402  
QY 4381 TCTCCCTAGTGAAGCAAGCCAGTGTCTTGTGAGAACCCACCGGCTGGCCGGGAAT 4440  
|||||  
Db 4403 TCTCCCTAGTGAAGCAAGCCAGTGTCTTGTGAGAACCCACCGGCTGGCCGGGAAT 4462  
QY 4441 TTTACAGCAAGCGGCTGTGGATTAATCTTGTGAAATTTCAATCTTCCCGGCT 4500  
|||||  
Db 4463 TTTACAGCAAGCGGCTGTGGATTAATCTTGTGAAATTTCAATCTTCCCGGCT 4522  
QY 4501 CTGTGTGAGCCCATCTGTGTATCTGTGTGTTTGGACCCCTAATGTCAAGTGGCT 4560  
|||||  
Db 4523 CTGTGTGAGCCCATCTGTGTATCTGTGTGTTTGGACCCCTAATGTCAAGTGGCT 4582  
QY 4561 GTAGGACATCCCGAGGTTTGGATGTGTGTGTAAGCAATGGAGGCTGTATTTGGCTGTGA 4620  
|||||  
Db 4583 GTAGGACATCCCGAGGTTTGGATGTGTGTGTAAGCAATGGAGGCTGTATTTGGCTGTGA 4642  
QY 4621 AGCTACATCCAGCCTTGGATTTAAGGAGGATTCACACCGGATTTACACTTTCCACT 4680  
|||||  
Db 4643 AGCTACATCCAGCCTTGGATTTAAGGAGGATTCACACCGGATTTACACTTTCCACT 4702  
QY 4681 CCTGTAGGATTTCTGTCCCTGGGCTGAATGAATATATTTTGGGTACAG 4740  
|||||







MADNGDLIDLPVEMEPYASPRYLBNFLTVDPYISMKCAPPEPHLEPHTPDPVLP  
SEAKPLGSGTGERPRIGGAGSGSKREFOALAGVNPQDFVWLTJYEREVLNLSLER  
LNLPIYLNKRYVYVWNSPKLPSEDLPLMPDVGPIIMVYFTENKSNLNNRFLPWNIEJETA  
LSTDDAHLMHDEIMGFVYVREARDRIVGFPRYAMDPHQSMYXSNYSYSELSMY  
LTGAAPFHXYALYSYVMPQALRDYVDEINCEIDYAMNVLVSHITRKPRIKYTSMT  
FRCPGPOALSHSDSHFHERHKCINFEVYVGMPLLYIQFRNDSVLEFKTRLPDHDXTK  
CFEFL

BASE COUNT 1261 a 1681 c 1694 g 1511 t  
ORIGIN

Query Match 99.2%; Score 6124.8; DB 9; Length 6147;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6137; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGGGGTCCCTGAGCTGGAGGCGGAGAGACCCCTGAGGTTCACTCTTCAAGAACT 60  
DB 9 GGGGGTCCCTGAGCTGGAGGCGGAGAGACCCCTGAGGTTCACTCTTCAAGAACT 68  
QY 61 CGTGTCTAGAGTGTATCTACACAAGTCAAGAGGAAGGAGGCTCTGAACACATGGC 120  
DB 69 CGTGTCTAGAGTGTATCTACACAAGTCAAGAGGAAGGAGGCTCTGAACACATGGC 128  
QY 121 CTGATTGTGGCAAGGACATCAAGAACTGGCATTTATTTCTGTCTAACCTATTACT 180  
DB 129 CTGATTGTGGCAAGGACATCAAGAACTGGCATTTATTTCTGTCTAACCTATTACT 188  
QY 181 GTATACCTGTGAATGACACTATGATATTTGTGTGCGCAAAAACCAAGAAAGAGC 240  
DB 189 GTATACCTGTGAATGACACTATGATATTTGTGTGCGCAAAAACCAAGAAAGAGC 248  
QY 241 TATGCAATTTGAAAAAGTGTGTCTGATTCAGAGGTGTTTCTGGGTTTCACTCATAGG 300  
DB 249 TATGCAATTTGAAAAAGTGTGTCTGATTCAGAGGTGTTTCTGGGTTTCACTCATAGG 308  
QY 301 TACCTCCCTCCCTTTCATCTCAAGCAAGAACTGGCACCTTTATCTGTGTGAATTAAGATTAA 360  
DB 309 TACCTCCCTCCCTTTCATCTCAAGCAAGAACTGGCACCTTTATCTGTGTGAATTAAGATTAA 368  
QY 361 GGACATGTTCTTGGTCAACAGCCAGAACTTAAAACTGCGGAATAGGGTCAAGACCA 420  
DB 369 GGACATGTTCTTGGTCAACAGCCAGAACTTAAAACTGCGGAATAGGGTCAAGACCA 428  
QY 421 TTTTACGTGCAAGTGAAGAAAAATGTAATGTTTATTTTGGGCGCTGTCTGGGAGC 480  
DB 429 TTTTACGTGCAAGTGAAGAAAAATGTAATGTTTATTTTGGGCGCTGTCTGGGAGC 488  
QY 481 ACACCTACTCTTCTGGAAGAGTGTCACTGAACAGAGATCGTTTGTGGAATAGCAACC 540  
DB 489 ACACCTACTCTTCTGGAAGAGTGTCACTGAACAGAGATCGTTTGTGGAATAGCAACC 548  
QY 541 ATGGTTATGCGAGTGAACCCGAGTGTCTGGGGGAGAGCTGACAGAGACTCATGACAG 600  
DB 549 ATGGTTATGCGAGTGAACCCGAGTGTCTGGGGGAGAGCTGACAGAGACTCATGACAG 608  
QY 601 GCTATACCATGCTGCGGAATGGGGGCGCGGGAAGAGAGTCAACCTGATCTGGGCT 660  
DB 609 GCTATACCATGCTGCGGAATGGGGGCGCGGGAAGAGTCAACCTGATCTGGGCT 668  
QY 661 GGTCCAAACCGCATCCGCTTACGTGCTCACTTCAAGCGCTTTGTATCACTCGTCTTC 720  
DB 669 GGTCCAAACCGCATCCGCTTACGTGCTCACTTCAAGCGCTTTGTATCACTCGTCTTC 728  
QY 721 TCCCGCTCATCGCCCACTTTACCTCAACACTCTGATAGAGCTGATGAGCGAGCGCAAGC 780  
DB 729 TCCCGCTCATCGCCCACTTTACCTCAACACTCTGATAGAGCTGATGAGCGAGCGCAAGC 788  
QY 781 GGAATTTTGGTCCCGGGGTGGGGAACGAGCTGTGCGAGGTGAAGCAAGCTGCTGATCT 840  
DB 789 GGAATTTTGGTCCCGGGGTGGGGAACGAGCTGTGCGAGGTGAAGCAAGCTGCTGATCT 848  
QY 841 GCCGATCCGGAGTGTGAGTGAAGACTCTCTCAGCTGAGGCGCAAGCGCCAAAGAGC 900  
DB 849 GCCGATCCGGAGTGTGAGTGAAGACTCTCTCAGCTGAGGCGCAAGCGCCAAAGAGC 908

DB 849 GCCGATCCGGAGTGTGAGTGAAGACTCTCTCAGCTGAGGCGCAAGCGCCAAAGAGC 908  
QY 901 TGAACACGAGATGCGCAAGCTGATGTGAAGATGGAAGCGCTGTGAAGAGCATTTAGA 960  
DB 909 TGAACACGAGATGCGCAAGCTGATGTGAAGATGGAAGCGCTGTGAAGAGCATTTAGA 968  
QY 961 ACCGACAGAGACCTCTCAGCTCAAGAACTGATGATCAGACAGACCGAGCATTTCTACA 1020  
DB 969 ACCGACAGAGACCTCTCAGCTCAAGAACTGATGATCAGACAGACCGAGCATTTCTACA 1028  
QY 1021 AGAAGCTCATGCGCCAGAAACAGCCAGCTGTCCCTGAGGCGCAAGCGCTGTGAAGAG 1080  
DB 1029 AGAAGCTCATGCGCCAGAAACAGCCAGCTGTCCCTGAGGCGCAAGCGCTGTGAAGAG 1088  
QY 1081 AGAAGCATGCGCGCTCTCTCCCTCCGAGAGCCACTGCGGGCTGCGCGCTACACACTGCT 1140  
DB 1089 AGAAGCATGCGCGCTCTCTCCCTCCGAGAGCCACTGCGGGCTGCGCGCTACACACTGCT 1148  
QY 1141 TTGATTAATCTGCTTCCCTCTCACTCTGAGGCTTCCCGGCTTACGCTTATGACAGTGAAC 1200  
DB 1149 TTGATTAATCTGCTTCCCTCTCACTCTGAGGCTTCCCGGCTTACGCTTATGACAGTGAAC 1208  
QY 1201 AGTTGTCTTTGGCAGCTACCTGATCCCTGGTCAAGCAGGCTTTTCAAGGAGACAGC 1260  
DB 1209 AGTTGTCTTTGGCAGCTACCTGATCCCTGGTCAAGCAGGCTTTTCAAGGAGACAGC 1268  
QY 1261 GAGCTAAGCTTATATTTACAGAAAAATGCAAGATGCGCTTACGTGATGATGAG 1320  
DB 1269 GAGCTAAGCTTATATTTACAGAAAAATGCAAGATGCGCTTACGTGATGATGAG 1328  
QY 1321 GAGAGATGAGAGGCGCGTGTGTGCGGCGCTGCTGAGCTGGAGAGAGTGTTCCTC 1380  
DB 1329 GAGAGATGAGAGGCGCGTGTGTGCGGCGCTGCTGAGCTGGAGAGAGTGTTCCTC 1388  
QY 1381 TGCCACATGCGCGGAGCGGATGAGACACAACATGTCATCATCAATCTGTACAGTAAGTCAG 1440  
DB 1389 TGCCACATGCGCGGAGCGGATGAGACACAACATGTCATCATCAATCTGTACAGTAAGTCAG 1448  
QY 1441 ATACACAGAACTTCTCTATAGCTGAGTACTGGGCGCGGCTGATGAGAGCGCCGCTCT 1500  
DB 1449 ATACACAGAACTTCTCTATAGCTGAGTACTGGGCGCGGCTGATGAGAGCGCCGCTCT 1508  
QY 1501 TCTACACTGTCCAGTACAGACCTGGCTTGTGACTGTGTGCTATCAACCGCTGGTCCATGCCA 1560  
DB 1509 TCTACACTGTCCAGTACAGACCTGGCTTGTGACTGTGTGCTATCAACCGCTGGTCCATGCCA 1568  
QY 1561 TGTGTGAGCCCACTTATGGAATCCACACAGATGCGCGGTGAAGCGGAAATATCTCT 1620  
DB 1569 TGTGTGAGCCCACTTATGGAATCCACACAGATGCGCGGTGAAGCGGAAATATCTCT 1628  
QY 1621 TCACCTTCCAGGGGAGAAATTTGAGTCTGTGAGGTCTACCTTTCAAGAGCGCCGCTCT 1680  
DB 1629 TCACCTTCCAGGGGAGAAATTTGAGTCTGTGAGGTCTACCTTTCAAGAGCGCCGCTCT 1688  
QY 1681 TCGAAGAGAAATGAGGGGAGCCCTCCGCGACTACGATGACGCGGATATTGGCCACCC 1740  
DB 1689 TCGAAGAGAAATGAGGGGAGCCCTCCGCGACTACGATGACGCGGATATTGGCCACCC 1748  
QY 1741 TGAAGCGGTGCAAGAGCAAGCTGATCAGGCTCTGTGTGAATTCACCTGCAAAAAAC 1800  
DB 1749 TGAAGCGGTGCAAGAGCAAGCTGATCAGGCTCTGTGTGAATTCACCTGCAAAAAAC 1808  
QY 1801 AGCCCAACCCAGCCGCGGAGCTGATGAGGCACTGTGTGAGAGCGGGAAGACCCGCTGG 1860  
DB 1809 AGCCCAACCCAGCCGCGGAGCTGATGAGGCACTGTGTGAGAGCGGGAAGACCCGCTGG 1868  
QY 1861 AATGTGTGAAGCTTCCACCTTCCGCTCATCATTAACCCGCGGGAACCTCGCTTGGTTA 1920  
DB 1869 AATGTGTGAAGCTTCCACCTTCCGCTCATCATTAACCCGCGGGAACCTCGCTTGGTTA 1928  
QY 1921 TTTCTCTGGGTGTCAACACGCGCTTTGGAAGCCCTGGAAGTCTGCTCCGCGGTGG 1980  
DB 1929 TTTCTCTGGGTGTCAACACGCGCTTTGGAAGCCCTGGAAGTCTGCTCCGCGGTGG 1988

QY 1981 TCGTGGGGAGGAGCTCAGCTTCCCTACCAGACATGCTGCAGTGGAAACGAGCGGCC 2040  
| | | | |  
Db 1989 TCGTGGGGAGGAGCTCAGCTTCCCTACCAGACATGCTGCAGTGGAAACGAGCGGCC 2048  
QY 2041 TGGTGGTGCAGAAAGCTCGCTGTTACCGAGGTTTCATTTCCGTCAGAGAGCTTCCGATA 2100  
| | | | |  
Db 2049 TGGTGGTGCAGAAAGCTCGCTGTTACCGAGGTTTCATTTCCGTCAGAGAGCTTCCGATA 2108  
QY 2101 GTGACCTCTGCTGATAGAGCGGCAAGCGCGCTTCTCTGGAGAGCTTACTTCTCCACTG 2160  
| | | | |  
Db 2109 GTGACCTCTGCTGATAGAGCGGCAAGCGCGCTTCTCTGGAGAGCTTACTTCTCCACTG 2168  
QY 2161 CTGACAGTATTTTAAATACCGTCTGCTGATATGATAGAGCTGCATCCAGATCCAGCCG 2220  
| | | | |  
Db 2169 CTGACAGTATTTTAAATACCGTCTGCTGATATGATAGAGCTGCATCCAGATCCAGCCG 2228  
QY 2221 CTCGCCATCCGGGAGAGGGGGGAGCTGAGATCCCGACCGTTACAGGCAAGCGCGTGGAA 2280  
| | | | |  
Db 2229 CTCGCCATCCGGGAGAGGGGGGAGCTGAGATCCCGACCGTTACAGGCAAGCGCGTGGAA 2288  
QY 2281 CTGACCCCAACATGCTGACAAACGGGGACCTGGACCTGGGGCCAGTGAAGACGAGCCG 2340  
| | | | |  
Db 2289 CTGACCCCAACATGCTGACAAACGGGGACCTGGACCTGGGGCCAGTGAAGACGAGCCG 2348  
QY 2341 CCTACGCTTACCAGATACCTCCGCAATTTCACTCTGACTGACTGACTTATACCGCA 2400  
| | | | |  
Db 2349 CCTACGCTTACCAGATACCTCCGCAATTTCACTCTGACTGACTGACTTATACCGCA 2408  
QY 2401 GGTGGAAGTGTGCTCAGGGGCTTCCATCTTTCCCGCAACACTCCGCTTGAACCTGTG 2460  
| | | | |  
Db 2409 GGTGGAAGTGTGCTCAGGGGCTTCCATCTTTCCCGCAACACTCCGCTTGAACCTGTG 2468  
QY 2461 TGGCCTCAGAGGCAAAATCTTGGGCTCAGGAGCTGCTTCCGCTTATTTGATGATGAG 2520  
| | | | |  
Db 2469 TGGCCTCAGAGGCAAAATCTTGGGCTCAGGAGCTGCTTCCGCTTATTTGATGATGAG 2528  
QY 2521 CTGGGGGTTCTGGCAAGAAATTCAGGCAAGCTTGGAGCAATGTTCCCGAGACAGT 2580  
| | | | |  
Db 2529 CTGGGGGTTCTGGCAAGAAATTCAGGCAAGCTTGGAGCAATGTTCCCGAGACAGT 2588  
QY 2581 TCAAGGTGGATGTTGACTTATGAGCGGAGAGAGTGGTATGAATCTTTAGAGAGG 2640  
| | | | |  
Db 2589 TCAAGGTGGATGTTGACTTATGAGCGGAGAGAGTGGTATGAATCTTTAGAGAGG 2648  
QY 2641 TGAATGGCTTCCCTTACCTGAACAGAGTGTGTGTGTAATTCCTCCAGAGCTCCAT 2700  
| | | | |  
Db 2649 TGAATGGCTTCCCTTACCTGAACAGAGTGTGTGTGTAATTCCTCCAGAGCTCCAT 2708  
QY 2701 CAGAGACCTTCTGTGGCTGACATTTGGGTTCCCATGATGTTCCGTTACTGAGAGA 2760  
| | | | |  
Db 2709 CAGAGACCTTCTGTGGCTGACATTTGGGTTCCCATGATGTTCCGTTACTGAGAGA 2768  
QY 2761 ACAGTTTGAACAACGATTTCTTACCTGGAATGAATTTGAGAGAGGCAATCTTCCA 2820  
| | | | |  
Db 2769 ACAGTTTGAACAACGATTTCTTACCTGGAATGAATTTGAGAGAGGCAATCTTCCA 2828  
QY 2821 TTGATAGCAGATGCTACCTCCGCAATGAGAAATCATGTTGGGTTCCGGGTGTGAGAG 2880  
| | | | |  
Db 2829 TTGATAGCAGATGCTACCTCCGCAATGAGAAATCATGTTGGGTTCCGGGTGTGAGAG 2888  
QY 2881 AAGCTGGGAGCGGATCGTGGGCTTCCCTGGGCTTACCAAGATGGAGATCCCCATC 2940  
| | | | |  
Db 2889 AAGCTGGGAGCGGATCGTGGGCTTCCCTGGGCTTACCAAGATGGAGATCCCCATC 2948  
QY 2941 AGTCTGGGCTTCACTCACTCAACTACTCTGTGAGCTGTCCATGTGACAGAGTGTG 3000  
| | | | |  
Db 2949 AGTCTGGGCTTCACTCACTCAACTACTCTGTGAGCTGTCCATGTGACAGAGTGTG 3008  
QY 3001 CCTTCTTTCACAAGTATATGCTTACCTGATATTTATGATGATGCCAGGCGCATCGG 3060  
| | | | |  
Db 3009 CCTTCTTTCACAAGTATATGCTTACCTGATATTTATGATGATGCCAGGCGCATCGG 3068

QY 3061 ACATGSGATGATATCATCACTGAGAGACATTTGCCATGAATCTTCTGTCTCCACA 3120  
| | | | |  
Db 3069 ACATGSGATGATATCATCACTGAGAGACATTTGCCATGAATCTTCTGTCTCCACA 3128  
QY 3121 TCACCTGGAAGCCCCCATCAAGTACCTCAAGTGGAGATTTCCGATCCAGATGCC 3180  
| | | | |  
Db 3129 TCACCTGGAAGCCCCCATCAAGTACCTCAAGTGGAGATTTCCGATCCAGATGCC 3188  
QY 3181 CTCAGGCCCTGTCTCAATGATGATCCCACTTCCAGAGGCGGCAAGTGCATCACTTCT 3240  
| | | | |  
Db 3189 CTCAGGCCCTGTCTCAATGATGATCCCACTTCCAGAGGCGGCAAGTGCATCACTTCT 3248  
QY 3241 TGTGGAAGTGTACGCTACATGCCCCCTCTGTACACGAGTTCAAGGTGATTTCTGTC 3300  
| | | | |  
Db 3249 TGTGGAAGTGTACGCTACATGCCCCCTCTGTACACGAGTTCAAGGTGATTTCTGTC 3308  
QY 3301 TCTTCAAGACAGCCCTGCCCATGACAAAGCAAGGCTTCAAGTTCACTTAGGGGACG 3360  
| | | | |  
Db 3309 TCTTCAAGACAGCCCTGCCCATGACAAAGCAAGGCTTCAAGTTCACTTAGGGGACG 3368  
QY 3361 GCACGCTCTGGGAGAGAGATGAGCAGAGGAGAGATGCTCCCAAGGTTCTTAGGCA 3420  
| | | | |  
Db 3369 GCACGCTCTGGGAGAGAGATGAGCAGAGGAGAGATGCTCCCAAGGTTCTTAGGCA 3428  
QY 3421 TTGACAGACCTTGGGACATCTGCTGTGGTGGCCCAAGCCTCTGCTGGAAGGGGCA 3480  
| | | | |  
Db 3429 TTGACAGACCTTGGGACATCTGCTGTGGTGGCCCAAGCCTCTGCTGGAAGGGGCA 3488  
QY 3481 CAGAGAGTGTGGAAGAAACCGCTGCTTATCTTGAAGTCAAGCCACATCGGGCCGGAG 3540  
| | | | |  
Db 3489 CAGAGAGTGTGGAAGAAACCGCTGCTTATCTTGAAGTCAAGCCACATCGGGCCGGAG 3548  
QY 3541 CCCTGGGCGAGTCCCGGGGTTCCCAACAGAGGCACTGACTATAGCTTACACTGAG 3600  
| | | | |  
Db 3549 CCCTGGGCGAGTCCCGGGGTTCCCAACAGAGGCACTGACTATAGCTTACACTGAG 3608  
QY 3601 ACTGTGGCACTGTGAGAGTCACTACACCGTGTACGCCAGGACAGCTGTGCTG 3660  
| | | | |  
Db 3609 ACTGTGGCACTGTGAGAGTCACTACACCGTGTGTACGCCAGGACAGCTGTGCTG 3668  
QY 3661 GTTTTTACATTCATTAACAATATATATATTTTAAAGAGAAATTTAGATTTGCC 3720  
| | | | |  
Db 3669 GTTTTTACATTCATTAACAATATATATATTTTAAAGAGAAATTTAGATTTGCC 3728  
QY 3721 ATTCAAGGCTTATATATATATATGTTGTATATATATATATATATATATATATAT 3780  
| | | | |  
Db 3729 ATTCAAGGCTTATATATATATATGTTGTATATATATATATATATATATATATAT 3788  
QY 3781 TATATATTTTGGCTGGGAGAGTGTATTTGCCCTTTCTAAGGAGGAGCCGCGAGG 3840  
| | | | |  
Db 3789 TATATATTTTGGCTGGGAGAGTGTATTTGCCCTTTCTAAGGAGGAGCCGCGAGG 3848  
QY 3841 TCCCTTGTCTGATCTGAGGAGATGGGCTCTGCTTGTGTCATCGCTTATCTTA 3900  
| | | | |  
Db 3849 TCCCTTGTCTGATCTGAGGAGATGGGCTCTGCTTGTGTCATCGCTTATCTTA 3908  
QY 3901 AAGATCATCTCCATCTCCAGGCGCATGTGTGACAGCAACGAAAGGATGAAT 3960  
| | | | |  
Db 3909 AAGATCATCTCCATCTCCAGGCGCATGTGTGACAGCAACGAAAGGATGAAT 3968  
QY 3961 TGGCCCTCTTGGGGGCTGGAGCAAGGTCCTTACCCTTCTGTGTCAGTCCAGNA 4020  
| | | | |  
Db 3969 TGGCCCTCTTGGGGGCTGGAGCAAGGTCCTTACCCTTCTGTGTCAGTCCAGNA 4028  
QY 4021 CCTGTAACTCAATTCCTTCCAGTGAATCCCTGGAGAGGCTGACCTGTGGGCTGT 4080  
| | | | |  
Db 4029 CCTGTAACTCAATTCCTTCCAGTGAATCCCTGGAGAGGCTGACCTGTGGGCTGT 4088  
QY 4081 TCAGCTTCTGCTGTGGGCGAGCGATTTTGAAGATTTATCTTATAGCCAGGCTTGC 4140  
| | | | |  
Db 4089 TCAGCTTCTGCTGTGGGCGAGCGATTTTGAAGATTTATCTTATAGCCAGGCTTGC 4148  
QY 4141 TCCGTAATTAATCCGCTCTCCCATTTCTCTGTGTTAGAGAGATAGAGAGCAAG 4200

Db 4149 TCCGACTATATCCCTGCTCCCACTTTCTCTTTGTTGAGAGGAATGAGGAAGCAAG 4208  
Qy 4201 AGGAGAGAAATATAGGGGCTGAAAGAGCCACTCCAGATGGCTTTCTATCCGCTCTT 4260  
Db 4209 AGGAGAGAAATATAGGGGCTGAAAGAGCCACTCCAGATGGCTTTCTATCCGCTCTT 4268  
Qy 4261 CTGTTAAACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4320  
Db 4269 CTGTTAAACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4328  
Qy 4321 ACAGGAGATCAGAGGCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4380  
Db 4329 ACAGGAGATCAGAGGCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4388  
Qy 4381 TCTCCCTAGTATAGAGCAAGCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4440  
Db 4389 TCTCCCTAGTATAGAGCAAGCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4448  
Qy 4441 TTATACAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4500  
Db 4449 TTATACAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4508  
Qy 4501 CTGTTGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4560  
Db 4509 CTGTTGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4568  
Qy 4561 GTAGAGCTCCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4620  
Db 4569 GTAGAGCTCCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4628  
Qy 4621 AGCTCACATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4680  
Db 4629 AGCTCACATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4688  
Qy 4681 CCCGCTTAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4740  
Db 4689 CCCGCTTAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4748  
Qy 4741 GTGCACTAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4800  
Db 4749 GTGCACTAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4808  
Qy 4801 GTGTTCTTGAACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4860  
Db 4809 GTGTTCTTGAACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4868  
Qy 4861 AACAAAAGCTCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
Db 4869 AACAAAAGCTCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4928  
Qy 4921 GACCTGGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4980  
Db 4929 GACCTGGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4988  
Qy 4981 AGCTGGGCTCTCTTCAAGTACTGTTTGAAGCTGGGCTCTTGTGTAGTCCACCCAC 5040  
Db 4989 AGCTGGGCTCTCTTCAAGTACTGTTTGAAGCTGGGCTCTTGTGTAGTCCACCCAC 5048  
Qy 5041 CTGTAGGCTAGCTGGCTTAAAGGAAGCTCCCATTTGCAAGAGGAGCCGGCCGGCG 5100  
Db 5049 CTGTAGGCTAGCTGGCTTAAAGGAAGCTCCCATTTGCAAGAGGAGCCGGCCGGCG 5108  
Qy 5101 CCAGAGCTGTCTTCCAAAGGTTCCCGCCCAACCCAGATCAGCTGTAGCTCCCGC 5160  
Db 5109 CCAGAGCTGTCTTCCAAAGGTTCCCGCCCAACCCAGATCAGCTGTAGCTCCCGC 5168  
Qy 5161 TGTCTAGGAGCTGTGTATGTTCCAGAGTGGGGGTAGACGCCCTTCTCAGAACTT 5220  
Db 5169 TGTCTAGGAGCTGTGTATGTTCCAGAGTGGGGGTAGACGCCCTTCTCAGAACTT 5228  
Qy 5221 TCTAGTTGCCCTTACCTGACTCCTGACTGTATCTTTTGGAGTGGCTCTTCTCCT 5280

Db 5229 TCTAGTTGCCCTTACCTGACTCCTGACTTGTATCTTTTACGAGTACCTTCTCCT 5288  
Qy 5281 CGGGAGACCAAGAGT 5340  
Db 5289 CGGGAGACCAAGAGT 5348  
Qy 5341 TTTTAAATGTAGAGAACTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5400  
Db 5349 TTTTAAATGTAGAGAACTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5408  
Qy 5401 TGT 5460  
Db 5409 TGT 5468  
Qy 5461 GGGGCTCTCTCTTATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5520  
Db 5469 GGGGCTCTCTCTTATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5528  
Qy 5521 AAGGCTGAGT 5580  
Db 5529 AAGGCTGAGT 5588  
Qy 5581 CAACAGAGACAGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5640  
Db 5589 CAACAGAGACAGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5648  
Qy 5641 CTGTAGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5700  
Db 5649 CTGTAGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5708  
Qy 5701 GGTGCTTCTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5760  
Db 5709 GGTGCTTCTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5768  
Qy 5761 CCACAGACCCCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820  
Db 5769 CCACAGACCCCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5828  
Qy 5821 CGCAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5880  
Db 5829 CGCAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5888  
Qy 5881 TCTTACATTAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5940  
Db 5889 TCTTACATTAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5948  
Qy 5941 GTGCTCTCCGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6000  
Db 5949 GTGCTCTCCGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6008  
Qy 6001 AGGATGGGATCTTCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060  
Db 6009 AGGATGGGATCTTCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6068  
Qy 6061 GAAGCTTTGATTCATAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6120  
Db 6068 GAAGCTTTGATTCATAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6128  
Qy 6121 GTCTTACATTTGATTTTGT 6140  
Db 6128 GTCTTACATTTGATTTTGT 6147

RESULT 5  
AF029231 5854 bp mRNA linear PRI 05-JAN-1999  
LOCUS AF029231 Homo sapiens EXT homolog mRNA, complete cds.  
DEFINITION AF029231  
ACCESSION AF029231.1 GI:4103883  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS	TITLE JOURNAL REFERENCE AUTHORS	TITLE JOURNAL REFERENCE AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 5854) Deng, H.-X., Xu, L., Xia, J.-H., Fan, C., Pan, Q., Liu, C.-Y. and Ruan, Q.-G.	Molecular cloning of a candidate gene for multiple exostoses unpublished 2 (bases 1 to 5854) Deng, H.-X., Xu, L., Xia, J.-H., Fan, C., Pan, Q., Liu, C.-Y. and Ruan, Q.-G.	Direct Submission Submitted (07-OCT-1997) National Laboratory of Medical Genetics, Hunan Medical University, 88 Xiangya Road, Changsha, Hunan 410078, P. R. of China

FEATURES	Location/Qualifiers
source	1. .5854

CDS

/noise+candidate gene for multiple exostoses"  
 /codon\_start+1  
 /product+*"EXT homolog 1"*  
 /protein\_id+*"AA01087.1"*  
 /db\_xref+*"GI:103884"*  
 /translation+*"MEKRYITLWRGANGNGGOTCMRLMSNRILRLTWLSTLFLVLFPPF  
 LIAHYLTTLDEADEAKRTITFGRVAVNCELVKHYHLDCIRINSEVSEELLDLLEKKRQD  
 LNSIATIKNTIEKCKSSINADODLQKNTYISOTESHYKCKFMAGNOKSLDPLRLR  
 PEKDGAPPPKATRGRLINCDCYRSCLTGTFGPPVYVDSQEFKQNSLPDLPVQKFA  
 NATRANVYVATENDALICGLVILVGGMOFPVLLRAPELEKQLYSLPHRTGGHHVITL  
 NLSKRPDQNLVYVSGRAMVAGNMFYTYVQPGDVLVSLVLAHSEPNEMELPROD  
 VPKRTILTFPGGKIKESLSSLSQAEKREEMEGEPDADYDRIATLAKVQSDQD  
 QVLEPFCNKDQSLPTPEALQSGEDDELTKLSTLRLITPDPLDPLVSSCAIRP  
 LPBALVGNAPVLYGEVQVLYPDQMLQNDMAALVYKPKPTVTHLLSLSDSLMLM  
 RQGGFLDELYTFSTADILFTVYLAIRTRITQDPAEIRREAAEELPHSGAAGTDVPL  
 MADQDLDELYVEETPEAPSRILNRFITLITQDPAEIRREAAEELPHSGAAGTDVPL  
 SEAPFLTSVGSGFRITGGGACGSEQOALAGVGNVPEQETVMYLLREVEELVMSLEL  
 NGNPLVFNKVVGNPRTSGDSELDLMPDLPVAVPRVETKNSLNLRPLMNEITELSL  
 LLSIDDADLRHDELMETFGFRVWARADRVYGFNCEIDANFVLSHTTRPDKTSMYNT  
 LTGADFAHKRYVLYVSYVMQALRDVMEYINCEIDANFVLSHTTRPDKTSMYNT  
 FRGCGPQALSHDSHNEHNRKCNINFEVYKVGYPMLLYTGFRVDSVLEKRLPDHDKTR  
 EFKR"*

BASE COUNT	1204 a	1612 c	1610 g	1428 t
ORIGIN				

Query Match	94.58;	Score 5834.6;	DB 9;	Length 5854;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 5837; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0.

	332	GGCACCTTTTATTCGTTTGATTAAGATTAAAGACATGTTCTTTGGTCAACAGCCGAACCTT	391
Db	1	GGCACCTTTTATTCGTTTATGATTAAGATTTAAGACATGTGTTCTTTGGTCAACAGCCGAACCTT	60
Qy	392	AAAATCTCTGGAAATAGGGTCAGAGACATTTCAGCTGCAGCTAGAGAAAATGAATGTT	451
Db	61	AAAATCTCTGGAAATAGGGTCAGAGACATTTCAGCTGCAGCTAGAGAAAATGAATGTT	120
Qy	452	CATTTATTTTGGTGCTGTCTGGGAGACACAACTCTCTGGAAACGTTAGTAA	511
Db	121	CATTTATTTTGGTGCTGTCTGGGAGACACAACTCTCTGGAAACGTTAGTAA	180
Qy	512	ACAGAGATCGTTTTGTGGAATAGCAACCACTGTTATGGGAGAGACCCGACGATCTG	571
Db	161	ACAGAGATCGTTTTGTGGAATAGCAACCACTGTTATGGGAGAGACCCGACGATCTG	240
Qy	572	GGGGGCAAGCTTCAGAGACTCATGACAGGCTATACATGCTGCGGAATGGGGGCGCGG	631
Db	241	GGGGGCAAGCTTCAGAGACTCATGACAGGCTATACATGCTGCGGAATGGGGGCGCGG	300
Qy	632	GAACGGAGGTACAGACTGCATGCTGCGCTGGTCCAACCGCATCCGCTCACTGTGCTCAG	691
Db	301	GAACGGAGGTACAGACTGCATGCTGCGCTGGTCCAACCGCATCCGCTCACTGTGCTCAG	360

Qy	692	CTCACAGCTCTTTTCATCCGGGCTCTTCCGGTCATCGCCACTATTACTACAC	751
Db	361	CTTACAGCTCTTTTCATCCGGCTTCTTCCGGTCATCGCCACTATTACTACAC	420
Qy	753	TCCTGATAGGCTATGAGGACGAGCAAGCGATTTTGTGCCGGGTGGGAGAGCT	811
Db	421	TCCTGATAGGCTATGAGGACGAGCAAGCGGATTTTGTGCCGGGTGGGAGAGCT	480
Qy	812	GTGCGAGGTGAAGACGTGTGATCTGTGCCGCATCCGGAGTCGGTAGTGAAGAGCT	871
Db	481	GTGCGAGGTGAAGACAGTGCTGATCTGTGCCGCATCCGGAGTCGGTAGTGAAGAGCT	540
Qy	872	CCTGACGTGGAAGGCCAAGGCCAAGAGCTGAACAGCGAGATTCGCCAAGCTGTA	931
Db	541	CCTGACGTGGAAGGCCAAGGCCAAGAGCTGAACAGCGAGATTCGCCAAGCTGTA	600
Qy	932	GATCCAAACCTGTGAAGAAAGACATTTGAAGAGCGCAAGCAGAGACCTCTCCAGCTCAAGA	991
Db	601	GATCCAAACCTGTGAAGAAAGACATTTGAAGAGCGCCAAAGCAGAGACCTCTCCAGCTCAAGA	660
Qy	992	TGTCATCAGCCAGACCGAGCATTCCTCAAGAGCTCATGCGCCAGAACCAAGCT	1051
Db	661	TGTCATCAGCCAGACCGAGCATTCCTCAAGAGCTCATGCGCCAGAACCAAGCT	720
Qy	1052	GTTCCTGCCCATCCGACTGCTCCCAAGAAAGAGATGCGCGGCTCCCTCCCCCAAGGC	1111
Db	721	GTTCCTGCCCATCCGACTGCTCCCAAGAAAGAGATGCGCGGCTCCCTCCCCCAAGGC	780
Qy	1112	CACCTGGGGGTCGCGGGCTACACAACCTCTTGAATTTCTCGTTGGCTCTCAGCTCTGG	1172
Db	781	CACCTGGGGGTCGCGGGCTACACAACCTCTTGAATTTCTCGTTGGCTCTCAGCTCTGG	840
Qy	1172	CTTCCCGGTCTACGCTCATATGACAGTGCACAGTTGTCTTTGGAGCTACCTGGATCCCTT	1231
Db	841	CTTCCCGGTCTACGCTCATATGACAGTGCACAGTTGTCTTTGGAGCTACCTGGATCCCTT	900
Qy	1232	GGTCAAGCAGGCTTTTCAAGCGGACAGACAGACAGCTTAAGTTACAGAAAATGACAA	1291
Db	901	GGTCAAGCAGGCTTTTCAAGCGGACAGACAGACAGCTTAAGTTACAGAAAATGACAA	960
Qy	1292	CATCCCTGCGCTTAAAGTGATACGTAAGTGGAGAGATGACAGAGCCGCTGGTCTCGGGC	1351
Db	961	CATCCCTGCGCTTAAAGTGATACGTAAGTGGAGAGATGACAGAGCCGCTGGTCTCGGGC	1020
Qy	1352	TGCTAGGCTGGGAAGCAGTTGTATTCCTGCCACACTGGCGGACGGATGGACACAACA	1411
Db	1021	TGCTAGGCTGGGAAGCAGTTGTATTCCTGCCACACTGGCGGACGGATGGACACAACA	1080
Qy	1412	TGTCATCATCAATCTGTACGTAAGTACATACAGAACCTTCTTATTAAGTCAGTAC	1472
Db	1081	TGTCATCATCAATCTGTACGTAAGTACATACAGAACCTTCTTATTAAGTCAGTAC	1140
Qy	1472	TGGCGGTGCATAGTGGGCGGACGTCACCTTCTACGTCAGTCCATACAGACCTGGGCTTGA	1531
Db	1141	TGGCGGTGCATAGTGGGCGGACGTCACCTTCTACGTCAGTCCATACAGACCTGGGCTTGA	1200
Qy	1532	CTTGGTCTATACACCGTGTGCATATGCCATATGTGTGAAGCCCAATCTATGAAATCCACAC	1591
Db	1201	CTTGGTCTATACACCGTGTGCATATGCCATATGTGTGAAGCCCAATCTATGAAATCCACAC	1260
Qy	1592	ACAGGTGCCGGTGAAGCGGAAATATCTCTTACCTTCACAGGCGGAGAAATTTGACTCTCT	1651
Db	1261	ACAGGTGCCGGTGAAGCGGAAATATCTCTTACCTTCACAGGCGGAGAAATTTGACTCTCT	1320
Qy	1652	GAGGTCTAGCCTTCAAGAGGCGCCGCTCTCTCGAAGAGAAATGGAAGGGGACCTCCCGC	1711
Db	1321	GAGGTCTAGCCTTCAAGAGGCGCCGCTCTCTCGAAGAGAAATGGAAGGGGACCTCCCGC	1380
Qy	1712	GCAGTACGATGACCGGATCATTTGCGACCTCGAAGGGGTCACAGACAGCAAGCTGGATGATCA	1772
Db	1381	GCAGTACGATGACCGGATCATTTGCGACCTCGAAGGGGTCACAGACAGCAAGCTGGATGATCA	1440
Qy	1772	GGTCTGTGTGAATTCACCTTGCAAAAAACGAGCCCAAAACCGAGCTCCCATCTGAGTGGGC	1831

Db 1441 GGTCTGGTGGAAATTCACCTGCAAAAAACACCCAAACCCAGCCCTGACCACTGATGGGC 1500  
QY 1832 ACTGTGGAGAGCGGGAGAGACCGCTTGGAAATGCTGAAGCTTCACCTTCGCCCTCAT 1891  
Db 1501 ACTGTGGAGAGCGGGAGAGACCGCTTGGAAATGCTGAAGCTTCACCTTCGCCCTCAT 1560  
QY 1892 CATTACCCCCGGGGACCCCTGCTGTTATTTCTCTGAGGTGTCACACAGCGCTTCGGA 1951  
Db 1561 CATTACCCCCGGGGACCCCTGCTGTTATTTCTCTGAGGTGTCACACAGCGCTTCGGA 1620  
QY 1952 AGCCCTTGGAGTGGTGGCCCTCCCGGTGCTGGGGGAGAGCTCCAGCTCCCTACCA 2011  
Db 1621 AGCCCTTGGAGTGGTGGCCCTCCCGGTGCTGGGGGAGAGCTCCAGCTCCCTACCA 1880  
QY 2012 GGCATGCTGCATGAGACAGAGCGGCCCTGGTGGTGGCCAAACCTTCGTTCACAGAT 2071  
Db 1681 GGCATGCTGCATGAGACAGAGCGGCCCTGGTGGTGGCCAAACCTTCGTTCACAGAT 1740  
QY 2072 TCATTTCCCTGTCAGAAAGCCCTCTCCGATAGTACCTCTGCTATGAGGCGGCAAGGCG 2131  
Db 1741 TCATTTCCCTGTCAGAAAGCCCTCTCCGATAGTACCTCTGCTATGAGGCGGCAAGGCG 1800  
QY 2132 CTTTCTCTGGGAGACTTACTTCTCCACTGCTACAGATATTTTAAATCCGTGCTGCTAT 2191  
Db 1801 CTTTCTCTGGGAGACTTACTTCTCCACTGCTACAGATATTTTAAATCCGTGCTGCTAT 1860  
QY 2192 GATTAGGACTGCATCCAGATCCAGCGCCCTCCATCCGGGAGAGAGCGGCGAGCTAGAT 2251  
Db 1861 GATTAGGACTGCATCCAGATCCAGCGCCCTCCATCCGGGAGAGAGCGGCGAGCTAGAT 1920  
QY 2252 CCCCCACCGTTCCAGCAAGCGCGCTGGAATGACCCCAACATGCTGACAAACGGGGAGCT 2311  
Db 1921 CCCCCACCGTTCCAGCAAGCGCGCTGGAATGACCCCAACATGCTGACAAACGGGGAGCT 1980  
QY 2312 GGACCTGGGGGCGAGGAGAGCGGCCCTAGCGCTCACCCAGATACCTCCGCAATTT 2371  
Db 1981 GGACCTGGGGGCGAGGAGAGCGGCCCTAGCGCTCACCCAGATACCTCCGCAATTT 2040  
QY 2372 CACTCTGACTGTCACACTTATTACCCAGCTGGAACGTGCTCCAGAGCGCTTTCATCT 2431  
Db 2041 CACTCTGACTGTCACACTTATTACCCAGCTGGAACGTGCTCCAGAGCGCTTTCATCT 2100  
QY 2432 TTTTCCCCACACTTCCCTTTTAACTGTTGCCCTCAGAGGCCAAATTTCTTGGGCTCAG 2491  
Db 2101 TTTTCCCCACACTTCCCTTTTAACTGTTGCCCTCAGAGGCCAAATTTCTTGGGCTCAG 2160  
QY 2492 GACTGCTTTCGGCTATTTGGTGGTGGAGCTGGGGTTCGCAAGGAATTTCAAGCAGC 2551  
Db 2161 GACTGCTTTCGGCTATTTGGTGGTGGAGCTGGGGTTCGCAAGGAATTTCAAGCAGC 2220  
Y 2552 GCTTGGAGGCAATGTTCCCGAGAGAGCTGACAGGTGGTGGATGTGACTTATATAGCGGGA 2611  
Db 2221 GCTTGGAGGCAATGTTCCCGAGAGAGCTGACAGGTGGTGGATGTGACTTATATAGCGGGA 2280  
QY 2612 GGAAGGCTTATGAACTCTTTAGAGAGCGTGAATGGCTCCCTTACCTGAACAAGGTGCT 2671  
Db 2281 GGAAGGCTTATGAACTCTTTAGAGAGCGTGAATGGCTCCCTTACCTGAACAAGGTGCT 2340  
QY 2672 GGTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTCTGTGGCTGACATTTGGCGT 2731  
Db 2341 GGTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTCTGTGGCTGACATTTGGCGT 2400  
QY 2732 TTCCATCATGCTGCTGCTACTGAGAGAAACAGTTTGAACAACGATTTCTTACCTGGAA 2791  
Db 2401 CCCCATCATGCTGCTGCTACTGAGAGAAACAGTTTGAACAACGATTTCTTACCTGGAA 2460  
QY 2792 TGAATGAGACAGAGCGCATCTGTCCATTTGATGACATGCTACCTCCGCAATGAGGA 2851  
Db 2461 TGAATGAGACAGAGCGCATCTGTCCATTTGATGACATGCTACCTCCGCAATGAGGA 2520  
QY 2852 AATCATTTGGGTTCCGGGTGTGAGAGAGCTCGGAGCCGATCGTGGGCTTCCCTGG 2911  
|||||

Db 2521 AATCATTTGGGTTCCGGGTGTGAGAGAGCTCGGAGCCGATCGTGGCTTCCCTGG 2580  
QY 2912 CCGTTACAGCATGGGACATCCCCCATCACTGCTGCTTACAACTCCAACTACTCTG 2971  
Db 2581 CCGTTACAGCATGGGACATCCCCCATCACTGCTGCTTACAACTCCAACTACTCTG 2640  
QY 2972 TGAGCTTCCATGCTGTGACAGAGTGCCTTCTTCAACAAGATTTATGCTTACGTGA 3031  
Db 2641 TGAGCTTCCATGCTGTGACAGAGTGCCTTCTTCAACAAGATTTATGCTTACGTGA 2700  
QY 3032 TTCTTATGTCATGCCCCAGGCGCATCCGGGACATGCTGATGAATATCACTCACTGAGGA 3091  
Db 2701 TTCTTATGTCATGCCCCAGGCGCATCCGGGACATGCTGATGAATATCACTGAGGA 2760  
QY 3092 CATTGCAATGAATCTCTTCTTCCCATCACTCGGAACCCCCCATCAAGGTGACCTC 3151  
Db 2761 CATTGCAATGAATCTCTTCTTCCCATCACTCGGAACCCCCCATCAAGGTGACCTC 2820  
QY 3152 ACGGTGGACATTCGATGCGCCAGAGATGCCCTGAGGCGCTGTCTCATGATGACTCCACTT 3211  
Db 2821 ACGGTGGACATTCGATGCGCCAGAGATGCCCTGAGGCGCTGTCTCATGATGACTCCACTT 2880  
QY 3212 CCACGAGCGGCAACAGTCATCACTTCTGCTGAAGGTGTACGGCTACATGCCCTCT 3271  
Db 2881 CCACGAGCGGCAACAGTCATCACTTCTGCTGAAGGTGTACGGCTACATGCCCTCT 2940  
QY 3272 GTACAGGAGTTCAGAGGTGATTTCTGTGCTTTCAGAACGCGCTGCCCATGACAAAC 3331  
Db 2941 GTACAGGAGTTCAGAGGTGATTTCTGTGCTTTCAGAACGCGCTGCCCATGACAAAC 3000  
QY 3332 CAAGTGTTCAGTTCATCTAGAGGCGAGCGGCTCTGGGGAGAGATGACAGAGG 3391  
Db 3001 CAAGTGTTCAGTTCATCTAGAGGCGAGCGGCTCTGGGGAGAGATGACAGAGG 3060  
QY 3392 AGGAAGATGCTCCCAAGGTTCTTACGATTCAGAGACCTTGGGCACTGCTGTGGGG 3451  
Db 3061 AGGAAGATGCTCCCAAGGTTCTTACGATTCAGAGACCTTGGGCACTGCTGTGGGG 3120  
QY 3452 TGGCCAGAGCTTGTGCTGGAGAGGGGAGAGAGAGAGTGGAGAAACCGCTCCCTTGA 3511  
Db 3121 TGGCCAGAGCTTGTGCTGGAGAGGGGAGAGAGAGAGTGGAGAAACCGCTCCCTTGA 3180  
QY 3512 TCTTGAAGTACAGCACACTGGGCTTGAGGCCCTGGGGAGTCCCGGGGTTCGCCACAC 3571  
Db 3181 TCTTGAAGTACAGCACACTGGGCTTGAGGCCCTGGGGAGTCCCGGGGTTCGCCACAC 3240  
QY 3572 AGGCACTGACTATGACTTACACTGAGACTGTGGGACTGCAAGATCACTACACAC 3631  
Db 3241 AGGCACTGACTATGACTTACACTGAGACTGTGGGACTGCAAGATCACTACACAC 3300  
QY 3632 GTTCGTACGCGCCAGGACAGCTGCTGTGGTTTTCATTTCACTTCAATAACAATATATGATT 3691  
Db 3301 GTTCGTACGCGCCAGGACAGCTGCTGTGGTTTTCATTTCACTTCAATAACAATATATGATT 3360  
QY 3692 ATTTAAAAAGAGAAAGTTTACATTTGCCATTCAGGCTTATTTATATATGCTGTGT 3751  
Db 3361 ATTTAAAAAGAGAAAGTTTACATTTGCCATTCAGGCTTATTTATATATGCTGTGT 3420  
QY 3752 ATATATAATCATGACACACTTTCATATATATATTTTGGCTGGGGAGTGTAGTTT 3811  
Db 3421 ATATATAATCATGACACACTTTCATATATATATTTTGGCTGGGGAGTGTAGTTT 3480  
QY 3812 TGCCCTTTGAAGGGGGGAGCGCGGAGGCTCTTGTGTTCTGTATTTCTGGCGAGATGGGT 3871  
Db 3481 TGCCCTTTGAAGGGGGGAGCGCGGAGGCTCTTGTGTTCTGTATTTCTGGCGAGATGGGT 3540  
QY 3872 CCTGGCTTGTGTACTGCTTATCTTAAAGATCATCTCCATCTCCAGCGGCAATC 3931  
Db 3541 CCTGGCTTGTGTACTGCTTATCTTAAAGATCATCTCCATCTCCAGCGGCAATC 3600  
QY 3932 TGTGTGAGCAACCAAGAAAGGATGAATTTGGCCCTTGGGGGCTTGGACAAAGTCTCT 3991  
Db 3601 TGTGTGAGCAACCAAGAAAGGATGAATTTGGCCCTTGGGGGCTTGGACAAAGTCTCT 3660  
|||||

OY 3992 TCCTTACCCCTTCTGTTGCCAGTCAGCAACCTGTAACTCACTCTCTTCCCACTGATC 4051  
|||||  
Db 3661 TCCTTACCCCTTCTGTTGCCAGTCAGCAACCTGTAACTCACTCTCTTCCCACTGATC 3720  
4052 CCTGGAGGCCCTGACCCCTGTTGGGCTGTTACAGTTTCCCTGCTGCTGGGGCCAGGATTTT 4111  
|||||  
Db 3721 CTTGGAGGCCCTGACCCCTGTTGGGCTGTTACAGTTTCCCTGCTGCTGGGGCCAGGATTTT 3780  
4112 TGAGAGATTATCTTTAGGCCAGGCTTCCCTCCACTTATCCCTGCTCCACTTCTCT 4171  
|||||  
Db 3781 TGAGAGATTATCTTTAGGCCAGGCTTCCCTCCACTTATCCCTGCTCCACTTCTCT 3840  
4172 CTTGTTTGAAGAGATATGAGAGAACAAAGAGTGAGAAAGATTAGGGCTGGAAGCCAC 4231  
3841 CTTGTTTGAAGAGATATGAGAGAACAAAGAGTGAGAAAGATTAGGGCTGGAAGCCAC 3900  
OY 4232 TCCGAGATGGCTCTTCTATCCCTGCTCTGTTGAAACACAGCTGCTGGGGCTCAGG 4291  
3901 TCCGAGATGGCTCTTCTATCCCTGCTCTGTTGAAACACAGCTGCTGGGGCTCAGG 3960  
4292 CGTTTCTGAAGTGCCTCTTCTGTTGAGATGACAGAGATCAGACGCTGACATCTGCT 4351  
|||||  
Db 3961 CGTTTCTGAAGTGCCTCTTCTGTTGAGATGACAGAGATCAGACGCTGACATCTGCT 4020  
4352 GGTCTGAAGTGTGTTGCAAGTCAAGCTCTCTCTCCCTAGTAGAGCAAGCCAGTCTCT 4411  
|||||  
Db 4021 GGTCTGAAGTGTGTTGCAAGTCAAGCTCTCTCTCCCTAGTAGAGCAAGCCAGTCTCT 4080  
4412 CGAGAGAACCCACCGCGCTGGCGGGAGTTTACAGAAAGCGCTGCTGGGATTAAT 4471  
4081 CGAGAGAACCCACCGCGCTGGCGGGAGTTTACAGAAAGCGCTGCTGGGATTAAT 4140  
OY 4472 CCTTGTGTAATTCACCTTCCCTCCGCTCTGCTGAGAGCCCATCTGTGTATCTGTG 4531  
4141 CCTTGTGTAATTCACCTTCCCTCCGCTCTGCTGAGAGCCCATCTGTGTATCTGTG 4280  
4532 GTTTTGGACCCCTTAATGTCAAGTCTGCTGTAGACTCCCGAGGTTTGGTATGCTAG 4591  
|||||  
Db 4201 GTTTTGGACCCCTTAATGTCAAGTCTGCTGTAGACTCCCGAGGTTTGGTATGCTAG 4260  
4592 AACAATGGAGGCTGTGTTGGTGTGTAGTCAATCCAGCCCTTGGAAATCAAGGGC 4651  
|||||  
Db 4261 AACAATGGAGGCTGTGTTGGTGTGTAGTCAATCCAGCCCTTGGAAATCAAGGGC 4320  
4652 ATTCAACAACCCGAGTTACCACTTCCACTCCCTTGAATTCCTTGTCCCTGGGCTGAA 4711  
|||||  
Db 4321 ATTCAACAACCCGAGTTACCACTTCCACTCCCTTGAATTCCTTGTCCCTGGGCTGAA 4380  
4712 ACTGAATTAAGCTAATTTTTGGGTCAAGGTGCAAGTAAAGGAGGCTGTAG 4771  
|||||  
Db 4381 ACTGAATTAAGCTAATTTTTGGGTCAAGGTGCAAGTAAAGGAGGCTGTAG 4440  
4772 TGGGATTTGTCAGGGAATTTACCCATGACGTGTTCTTGAAGCCCTACTTCTGGAGGG 4831  
4441 TGGGATTTGTCAGGGAATTTACCCATGACGTGTTCTTGAAGCCCTACTTCTGGAGGG 4500  
OY 4832 AGTTGACTCTGGAAGTTTCTTACGAACGTAAAGGCTAGGTTTGTCTGTATGCA 4891  
4501 AGTTGACTCTGGAAGTTTCTTACGAACGTAAAGGCTAGGTTTGTCTGTATGCA 4560  
OY 4892 CATGCCCTTAAGCCAGTCCGCTCTTCCCTAGACCTTGGCATCTGTCTTAATTTCTGG 4951  
4561 CATGCCCTTAAGCCAGTCCGCTCTTCCCTAGACCTTGGCATCTGTCTTAATTTCTGG 4620  
OY 4952 AATAGCTTCTCTGACCTGCTGACCTGTAACAGTGGGCTCTTCAAGTACGTTTGAAG 5011  
4621 AATAGCTTCTCTGACCTGCTGACCTGTAACAGTGGGCTCTTCAAGTACGTTTGAAG 4680  
OY 5012 CTGGGCTCTTCTGTAGCTCCCAACCACTGTAGGGCTAGCTGGCTTAAGGGAACTCT 5071  
4681 CTGGGCTCTTCTGTAGCTCCCAACCACTGTAGGGCTAGCTGGCTTAAGGGAACTCT 4740

OY 5072 CCCCATTTGGCAACCCGAGACCCGGCCGCCAGGACTGTGTTCCAAAGGTTCCCGCCC 5131  
|||||  
Db 4741 CCCCATTTGGCAACCCGAGACCCGGCCGCCAGGACTGTGTTCCAAAGGTTCCCGCCC 4800  
5132 CCAACCCGAGTACAGCTGTAGTCCCTGCTGAGGACGTGTATGTTCACAGAG 5191  
4801 CCAACCCGAGTACAGCTGTAGTCCCTGCTGAGGACGTGTATGTTCACAGAG 4860  
OY 5192 TGGGGGTCAAGCGCCCTTCTCAGAACTTCTAGTTGCCCTTACCTGACTCTGACTTC 5251  
4861 TGGGGGTCAAGCGCCCTTCTCAGAACTTCTAGTTGCCCTTACCTGACTCTGACTTC 4920  
OY 5252 TATTCCTTTTACAGATAGCTTCTCTCCCTGGGGAGCCAAAGATGTGTGTGGGGCT 5311  
4921 TATTCCTTTTACAGATAGCTTCTCTCCCTGGGGAGCCAAAGATGTGTGTGGGGCT 4980  
OY 5312 AATATTGGCTGATATTCATCTGTTCTTTTAAATGAGAACTCACAATGACTTC 5371  
4981 AATATTGGCTGATATTCATCTGTTCTTTTAAATGAGAACTCACAATGACTTC 5040  
OY 5372 AGTGGACTGCTGAGCGCGGGCCGCTGTGTGTGGGACCCCTTTAGCGGACTCAGT 5431  
5041 AGTGGACTGCTGAGCGCGGGCCGCTGTGTGTGGGACCCCTTTAGCGGACTCAGT 5100  
OY 5432 GAGCTGGGGCCCTGTGTGTGTGTGGAGCCAGGCTCTCCCTTAGTGGAGCAGTGTCTC 5491  
5101 GAGCTGGGGCCCTGTGTGTGTGTGGAGCCAGGCTCTCCCTTAGTGGAGCAGTGTCTC 5160  
OY 5492 GGGCCCGGAATGTACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5551  
5161 GGGCCCGGAATGTACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5220  
OY 5552 CGCAGGAGGAGGCTGT 5611  
5221 CGCAGGAGGAGGCTGT 5280  
OY 5612 GAACATGCTGT 5671  
5281 GAACATGCTGT 5340  
OY 5672 CAGCAAGTGTGCAAGT 5731  
5341 CAGCAAGTGTGCAAGT 5400  
OY 5732 TGTGACTGT 5791  
5401 TGTGACTGT 5460  
OY 5792 GGGGCGCCCTGGGCACTGT 5851  
5461 GGGGCGCCCTGGGCACTGT 5520  
OY 5852 TGGCGGCTGT 5911  
5521 TGGCGGCTGT 5580  
OY 5912 GGGCACTGT 5971  
5581 GGGCACTGT 5640  
OY 5972 AGGGCTCAACAGCTGGGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6031  
5641 AGGGCTCAACAGCTGGGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5700  
OY 6032 CAGAGTGGGAGCTCTAGTTGTGTGGGGGAAGCTTTGTATTCATGCCAGTCCATCA 6091  
5701 CAGAGTGGGAGCTCTAGTTGTGTGGGGGAAGCTTTGTATTCATGCCAGTCCATCA 5760  
OY 6092 CCCCAACCCCTTTTGT 6151  
5761 CCCCAACCCCTTTTGT 5820  
OY 6152 AATTAATGAGACTTTTAATC 6172



|||||

DB 5821 AATTAATGAGACTTACTC 5841

RESULT 6

LOCUS BC006363

DEFINITION Homo sapiens, exostoses (multiple)-like 3, clone MGC:12750 IMAGE:4131101, mRNA, complete cds.

ACCESSION BC006363

VERSION BC006363.1 GI:13623512

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3241)

REFERENCE 1 (bases 1 to 3241)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland.

Web site: [http://www.nisc.nih.gov/nisc\\_mgcenhgti.nih.gov](http://www.nisc.nih.gov/nisc_mgcenhgti.nih.gov)

Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantrilop, S., Thomas, P.J., Tlonsom, E.E., Touchman, D.W., Tsurgoun, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 17 Row: 1 Column: 10

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4103883.

location/Qualifiers

1. 3241

/organism="Homo sapiens"

/db\_xref="locustid:2137"

/db\_xref="taxon:9606"

/clone="MGC:12750 IMAGE:4131101"

/tissue="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

122..2881

/codon\_start=1

/product="exostoses (multiple)-like 3"

/protein\_id="AA06363.1"

/db\_xref="gi:13623513"

/translation="MTGYMLRKGAGNGGCGTCMLRNSRIRLITWSTFLVILVFP LTAHYLTLLDEADGAKRIFGRVNGNEICEVKAHVIDLCRIRSEVSEELIOLEAKOE LNSLEKLAKLKEACKRSIENAKODLQENKISQJESHSYKEIMANOQKLSPIELK PERKDAGLPKPRATRCRLHNCFDYSRCPLTGFPYVYSDQFVGSYLDPLVKQAF QATARANYVTENADIALCYLILVGMQEPVYLRAPELEKOLSLPHMRTDGHNYIT NSRSKDQNLILYVSTGRAMAOSTFYVOYRPGDLYVSPVHAMSPPNMEIPPO VPKRYLFTPGGKTESRSSLOEAKRSFEEMEGDPADYDRIITATLKAVDSKLD OYLVEPTCKNQKPSLPTFEMALCGERDELKLTSTLKITTPGDPRLVISGCAVTR LEAELEVGAVPVLEQVOLPYQDMLQNEALVYKPEVTEVHFLRLSISDLIAM RRGQFLMETVSTADISINTVIAMIRTIQIPAPVIREAAEIPDRHSKAGADPN MADNGDLIDGIVETEPYVAPRYLNFILITVTDYFRSMNCAAGCPHLPHPTFPDPLP

FEATURES

source

CDS

SEANFLSGTGRPDIGGAGGSGKFEQAALGNVPRQFTVVMLTREEVIANSLER  
LNGLDYLNKVVVWVNSPKLPSEDLMPDIPVIMVREKSNLNRFLPNEIEBAI  
LSIDDDALRHEDEIMFGRVREARDRIVGPRYHAMDIPOHSIMNSNSCETLSW  
LTGAFFHKYAYLYXSYVMPOAIRDMNDEYINCEDIAMNLFVSHITRPPYKTVSRRT  
FRCPGCPALSHSDSHFRHKKCNFVKYIGVIMPLLTORVDSVLFKTRLPDRTK  
CFKFI"

BASE COUNT 702 a 925 c 898 g 716 t

ORIGIN

Query Match 52.1%; Score 3217; DB 9; Length 3241;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 3220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 482 CACTACTCTTCTGGAAACGTGTCAGTGAACAGAGATCGTTTGTGAATAGCAACCA 541  
DB 10 CACTACTCTTCTGGAAACGTGTCAGTGAACAGAGATCGTTTGTGAATAGCAACCA 69

QY 542 TGGTTATGGGAGTGAACCCGAGAGTATCTGGGGGCGAGGACTCATGACAG 601  
DB 70 TGGTTATGGGAGTGAACCCGAGAGTATCTGGGGGCGAGGACTCATGACAG 129

QY 602 CTATACCATGCTGGGGAATGGGGGCGGCGAGAGAGTCAATGCTGCGCTG 661  
DB 130 CTATACCATGCTGGGGAATGGGGGCGGCGAGAGAGTCAATGCTGCGCTG 189

QY 662 GTCCAAACCCGATCCGCTCAGCTGAGCTCAGCTGCTTTGTGATCCTGGCTCTT 721  
DB 190 GTCCAAACCCGATCCGCTCAGCTGAGCTCAGCTGCTTTGTGATCCTGGCTCTT 249

QY 722 CCCGCTCATGCCCACTATTACCTACCATCTGTGATGAGGCTGATGAGGCAAGCG 781  
DB 250 CCCGCTCATGCCCACTATTACCTACCATCTGTGATGAGGCTGATGAGGCAAGCG 309

QY 782 GATTTTGGTCCCGGGGGGGAACGAGCTGTGGAGGTGAAGACGCTGATCTGTG 841  
DB 310 GATTTTGGTCCCGGGGGGGAACGAGCTGTGGAGGTGAAGACGCTGATCTGTG 369

QY 842 CCGCATCCGAGTGGTGAAGAGCTCTCAGCTGAGGCGCAAGCGCAAGAGCT 901  
DB 370 CCGCATCCGAGTGGTGAAGAGCTCTCAGCTGAGGCGCAAGCGCAAGAGCT 429

QY 902 GAACAGGAGATGCCCAAGCTGAATCTGAGATGGAAGCTGTGAAGAGCATTTGAA 961  
DB 430 GAACAGGAGATGCCCAAGCTGAATCTGAGATGGAAGCTGTGAAGAGCATTTGAA 489

QY 962 CGCCACGAGAGCTGCTCCAGCTCAAGATGATATAGCAGCAGAGATTCCTTCAA 1021  
DB 490 CGCCACGAGAGCTGCTCCAGCTCAAGATGATATAGCAGCAGAGATTCCTTCAA 549

QY 1022 GGAGCTCATGGCCAGAACAGCCCAAGCTGCTCCATCCATCGATGCTCCAGAGAA 1081  
DB 550 GGAGCTCATGGCCAGAACAGCCCAAGCTGCTCCATCCATCGATGCTCCAGAGAA 609

QY 1082 GGAGCATGCCGCTCCTCCCGGAAAGGCACTCGGGGCTGCCGGCTACAACTGCTT 1141  
DB 610 GGAGCATGCCGCTCCTCCCGGAAAGGCACTCGGGGCTGCCGGCTACAACTGCTT 669

QY 1142 TGAATATTCGTTGGCCCTGTCACCTGCTGCTCCGCGCTAGCTCATACAGTACCA 1201  
DB 670 TGAATATTCGTTGGCCCTGTCACCTGCTGCTCCGCGCTAGCTCATACAGTACCA 729

QY 1202 GTTGTCTTTGGAGCTACCTGATCCTTGTCAAGCAGAGCTTTTCAGGCGACAGCAG 1261  
DB 730 GTTGTCTTTGGAGCTACCTGATCCTTGTCAAGCAGAGCTTTTCAGGCGACAGCAG 789

QY 1262 AGCTAACGTTTATGTTACAGAAATGACATCGCCTGCTTTACGTGATTAAGTGGG 1321  
DB 790 AGCTAACGTTTATGTTACAGAAATGACATCGCCTGCTTTACGTGATTAAGTGGG 849

QY 1322 AGAGATGACGAGGACCGGTGCTGGGGGCTGCTGAGCTGAGAACAGATGATTCCT 1381  
DB 850 AGAGATGACGAGGACCGGTGCTGGGGGCTGCTGAGCTGAGAACAGATGATTCCT 909



QY 1382 GCCACACTGGCGGAGCGATGAGACACAACCAATGTCATCATCTGTACAGTAAGTCAGA 1441  
| | | | |  
Db 910 GCCACACTGGCGGAGCGATGAGACACAACCAATGTCATCATCTGTACAGTAAGTCAGA 969  
| | | | |  
QY 1442 TACACAGAACCTTCTCTATTAACGTCACTACTGCGCGTCCAGTGTGGCCCACTCACCTT 1501  
| | | | |  
Db 970 TACACAGAACCTTCTCTATTAACGTCACTACTGCGCGTCCAGTGTGGCCCACTCACCTT 1029  
| | | | |  
QY 1502 CTACACTGTCCAGTACAGACCTGCTTGAAGTGTGATACACGGTGGTGCATGTCAT 1561  
| | | | |  
Db 1030 CTACACTGTCCAGTACAGACCTGCTTGAAGTGTGATACACGGTGGTGCATGTCAT 1089  
| | | | |  
QY 1562 GTCTGAGCCCCAATCTCATATGGAATCCACACACAGGTGCCGTGAAGCGGAATATCTCTT 1621  
| | | | |  
Db 1090 GTCTGAGCCCCAATCTCATATGGAATCCACACACAGGTGCCGTGAAGCGGAATATCTCTT 1149  
| | | | |  
QY 1622 CACCTTCCAGGCGGACGAAGATGAGTCTGTAGAGTCTAGCTTACAGAGCGCCCTCTCTT 1681  
| | | | |  
Db 1150 CACCTTCCAGGCGGACGAAGATGAGTCTGTAGAGTCTAGCTTACAGAGCGCCCTCTCTT 1209  
| | | | |  
QY 1682 CGAAGAGGAATGAGAGGCGGACCTCCCGCGGACCTAGATACCGGATCATTTGCGACCTT 1741  
| | | | |  
Db 1210 CGAAGAGGAATGAGAGGCGGACCTCCCGCGGACCTAGATACCGGATCATTTGCGACCTT 1289  
| | | | |  
QY 1742 GAAGCGGCTGACAGAGACAGCAAGCTGATCAGTCTGGTGAATTCACCTGCAAAAACA 1801  
| | | | |  
Db 1270 GAAGCGGCTGACAGAGACAGCAAGCTGATCAGTCTGGTGAATTCACCTGCAAAAACA 1329  
| | | | |  
QY 1802 GCCCAAAACCCAGCTGCCGACTGAGTGGGACACTGTGTGAGAGCGGAGACCCCTTGA 1861  
| | | | |  
Db 1330 GCCCAAAACCCAGCTGCCGACTGAGTGGGACACTGTGTGAGAGCGGAGACCCCTTGA 1389  
| | | | |  
QY 1862 ATGCTGAAGCTTCCACCTTCCGCTCATCATTTACCCCGGGGACCTGCTGGTGTAT 1921  
| | | | |  
Db 1390 ATGCTGAAGCTTCCACCTTCCGCTCATCATTTACCCCGGGGACCTGCTGGTGTAT 1449  
| | | | |  
QY 1922 TTCTCTGGGCTGTGCAACACGCTCTTTCGAAGCCCTGGAATCGGTGCCCTCCGCTGCT 1981  
| | | | |  
Db 1450 TTCTCTGGGCTGTGCAACACGCTCTTTCGAAGCCCTGGAATCGGTGCCCTCCGCTGCT 1509  
| | | | |  
QY 1982 GCTGGGAGAGAGTCCAGCTTCCCTTACAGAGACATGCTGAGTGGAAACGAGGGGCCCT 2041  
| | | | |  
Db 1510 GCTGGGAGAGAGTCCAGCTTCCCTTACAGAGACATGCTGAGTGGAAACGAGGGGCCCT 1569  
| | | | |  
QY 2042 GGTGGTGCACAAAGCTCGTGTACAGAGGTTCAATTCCTGCTCAGAAACCTCTCGATAG 2101  
| | | | |  
Db 1570 GGTGGTGCACAAAGCTCGTGTACAGAGGTTCAATTCCTGCTCAGAAACCTCTCGATAG 1629  
| | | | |  
QY 2102 TGACCTCTGCTATGAGGCGGCAAGGCGCTTCTCTGGGAGACTTACTTCTCCACTGC 2161  
| | | | |  
Db 1630 TGACCTCTGCTATGAGGCGGCAAGGCGCTTCTCTGGGAGACTTACTTCTCCACTGC 1689  
| | | | |  
QY 2162 TGACAGTATTTTAAATACCGTGTGCTGCTATGATTAAGACTGCAATCCAGATCCAGCCGC 2221  
| | | | |  
Db 1690 TGACAGTATTTTAAATACCGTGTGCTGCTATGATTAAGACTGCAATCCAGATCCAGCCGC 1749  
| | | | |  
QY 2222 TCCCATCCGGGAGAGGCGGAGCTGAGATCCCAACGTTTCAGGCAAGGCGGCTGGAAC 2281  
| | | | |  
Db 1750 TCCCATCCGGGAGAGGCGGAGCTGAGATCCCAACGTTTCAGGCAAGGCGGCTGGAAC 1809  
| | | | |  
QY 2282 TGACCCCAACATGCTGACAAAGGCGGAGCTGAGCTTGGGCGAGTGGAGAGGAGCCGC 2341  
| | | | |  
Db 1810 TGACCCCAACATGCTGACAAAGGCGGAGCTGAGCTTGGGCGAGTGGAGAGGAGCCGC 1869  
| | | | |  
QY 2342 CTAGCGCTCACCACAGATACCTCCGCAATTTCACTGTCACTGACTTCTTACCCGAG 2401  
| | | | |  
Db 1870 CTAGCGCTCACCACAGATACCTCCGCAATTTCACTGTCACTGACTTCTTACCCGAG 1929  
| | | | |  
QY 2402 CTGGAAGTGTCTCAGGCGCTTTCACATCTTTCACCACTCCCTTGAACCTGTGT 2461  
| | | | |  
Db 1930 CTGGAAGTGTCTCAGGCGCTTTCACATCTTTCACCACTCCCTTGAACCTGTGT 1989  
| | | | |

QY 2462 GCCCTCAGAGCCCAATTTCTTGGGCTCAGGAGCTGCTTTCGGCTTATTTGTGTGAGC 2521  
| | | | |  
Db 1990 GCCCTCAGAGCCCAATTTCTTGGGCTCAGGAGCTGCTTTCGGCTTATTTGTGTGAGC 2049  
| | | | |  
QY 2522 TGGGGGTTTGGCAAGGAATTTTCAGGAGCGCTTGGAGGCAATGTTCCCGAGAGCATT 2581  
| | | | |  
Db 2050 TGGGGGTTTGGCAAGGAATTTTCAGGAGCGCTTGGAGGCAATGTTCCCGAGAGCATT 2109  
| | | | |  
QY 2582 CACGGTGTATGTTGACTTATGAGCGGAGGAAGTCTTATGAACCTTTAGAGAGCT 2641  
| | | | |  
Db 2110 CACGGTGTATGTTGACTTATGAGCGGAGGAAGTCTTATGAACCTTTAGAGAGCT 2169  
| | | | |  
QY 2642 GAATGCGCTCCCTTACTGAAACAAGTCTGTGTGTGGATTTCTCCAACTGCTCATC 2701  
| | | | |  
Db 2170 GAATGCGCTCCCTTACTGAAACAAGTCTGTGTGTGGATTTCTCCAACTGCTCATC 2229  
| | | | |  
QY 2702 AGAGAGCTTCTGTGGGCTGATGAGTGGCGTGTCCATCATGATGATGCTGCTGAGAA 2761  
| | | | |  
Db 2230 AGAGAGCTTCTGTGGGCTGATGAGTGGCGTGTCCATCATGATGATGCTGAGAA 2289  
| | | | |  
QY 2762 CAGTTTGAACAACCGATTTCTTACCTGGAATGAAATGAGACAGAGGCCATCTGTCCAT 2821  
| | | | |  
Db 2290 CAGTTTGAACAACCGATTTCTTACCTGGAATGAAATGAGACAGAGGCCATCTGTCCAT 2349  
| | | | |  
QY 2822 TGATGACGATGCTCACTCCGCGCATGACGAATATGTTTGGGTTCCGGGTGTGAGAGA 2881  
| | | | |  
Db 2350 TGATGACGATGCTCACTCCGCGCATGACGAATATGTTTGGGTTCCGGGTGTGAGAGA 2409  
| | | | |  
QY 2882 AGCTCGGGAGCCGATGCTGGGGCTTCCCGGCGTGTACACAGCATGGGACATCCCCCATCA 2941  
| | | | |  
Db 2410 AGCTCGGGAGCCGATGCTGGGGCTTCCCGGCGTGTACACAGCATGGGACATCCCCCATCA 2469  
| | | | |  
QY 2942 GTCTGCGCTCACAACCTCAACACTCTCTGTGAGCTGTCACTGTGCTGACAGGTGTGC 3001  
| | | | |  
Db 2470 GTCTGCGCTCACAACCTCAACACTCTCTGTGAGCTGTCACTGTGCTGACAGGTGTGC 2529  
| | | | |  
QY 3002 CTCTTTTCAACAATTAATGCTTACCTGTATTTATGTATGTCGCCAGGCCATCCGGA 3061  
| | | | |  
Db 2530 CTCTTTTCAACAATTAATGCTTACCTGTATTTATGTATGTCGCCAGGCCATCCGGA 2589  
| | | | |  
QY 3062 CATGCTGATGATATCAATCAACCTGTGAGGACATGGCATGAATCTCTGTCTCCCAT 3121  
| | | | |  
Db 2590 CATGCTGATGATATCAATCAACCTGTGAGGACATGGCATGAATCTCTGTCTCCCAT 2649  
| | | | |  
QY 3122 CACTCGGAAGCCGCCCATCAAGGTGACCTCAAGGTGACATTTCCAGATGCCAGATGCC 3181  
| | | | |  
Db 2650 CACTCGGAAGCCGCCCATCAAGGTGACCTCAAGGTGACATTTCCAGATGCCAGATGCC 2709  
| | | | |  
QY 3182 TCAGGCCCTGTCTCATGATGACTCCCACTTCCAGAGCGGACAAATGATCAACTTCTT 3241  
| | | | |  
Db 2710 TCAGGCCCTGTCTCATGATGACTCCCACTTCCAGAGCGGACAAATGATCAACTTCTT 2769  
| | | | |  
QY 3242 CGTGAAGTGTAGGCTACATAGTCCCTCTGTACAGGAGTTCAAGGAGTTCTGTGCT 3301  
| | | | |  
Db 2770 CGTGAAGTGTAGGCTACATAGTCCCTCTGTACAGGAGTTCAAGGAGTTCTGTGCT 2829  
| | | | |  
QY 3302 CTTCAAGACAGCCCTGCCCCATGACAAAGACCAAGTGTCTTCAAGTTCAATCTGAGGAGC 3361  
| | | | |  
Db 2830 CTTCAAGACAGCCCTGCCCCATGACAAAGACCAAGTGTCTTCAAGTTCAATCTGAGGAGC 2889  
| | | | |  
QY 3362 CACGGTCTGGGAGAGAGATGAGCAGAGGAGGAAGATGCTCCCAAGTCTTCAAGCAT 3421  
| | | | |  
Db 2890 CACGGTCTGGGAGAGAGATGAGCAGAGGAGGAAGATGCTCCCAAGTCTTCAAGCAT 2949  
| | | | |  
QY 3422 TGCAGAGCTTGGGCAATCTGCTGTGGGCGGAGGCTGCTGCTGGAAGGAGGAGC 3481  
| | | | |  
Db 2950 TGCAGAGCTTGGGCAATCTGCTGTGGGCGGAGGCTGCTGCTGGAAGGAGGAGC 3009  
| | | | |  
QY 3482 AGGAGAGTGGAGAAACCGCTGCTTATCTTGAAGTCAAGGCACTGAGGCTGAGAGC 3541  
| | | | |  
Db 3010 AGGAGAGTGGAGAAACCGCTGCTTATCTTGAAGTCAAGGCACTGAGGCTGAGAGC 3069  
| | | | |  
QY 3542 CTTGGGCGGAGTCCCGGGTTCCCAACAGAGGCACTGACTGATAGCTTACACTGAGGA 3601  
| | | | |



QY	4826	ACGCCAGCCTTGGAATCTAAAGGGGATCTCAACACCCGAGTTACACTTTCACATCCCTCG	4885
Db	66338	ACATCCAGCCCTTGGAATCTAAAGGGGATCTCAACACCCGAGTTACACTTTCACATCCCTCG	6627
QY	4886	CTTTGGAATCTTGTTCCCTGGGCGTAAACCTGAATTAAGCTAATTTTGTGGGTCAGCGTGGC	4745
Db	66278	CTTAAGGAATTTGTTCCCTGGGCGTAAACCTGAATTAAGCTAATTTTGTGGGTCAGCGTGGC	6621
QY	4746	AGTAGGGAACTAGGAGGGGTGTAGTGGCAATTTGTCAAGGATTTTAGCCATGACGTGT	4805
Db	66218	AGTAGGGAACTAGGAGGGGTGTAGTGGCAATTTGTCAAGGATTTTAGCCATGACGTGT	6615
QY	4806	TCTTGAACCTACTTTCTGGAAGTGGAGTGTGACTGTGGAAAGTTTCTTAGCACTGAACAA	4865
Db	66158	TCTTGAACCTACTTTCTGGAAGTGGAGTGTGACTGTGGAAAGTTTCTTAGCACTGAACAA	6609
QY	4866	AAGCTCAGGTTTCTCGGGGATGCACATGCCCTTAAGCAGTCCGTTCCCTAGACCT	4925
Db	66098	AAGCTCAGGTTTCTCGGGGATGCACATGCCCTTAAGCAGTCCGTTCCCTAGACCT	6603
QY	4926	TGGCATTCTGTGCTCTAATTTCTTGGGAATGCTTCCCTGACCTGGCCTTACACAGTG	4985
Db	66038	TGGCATTCTGTGCTCTAATTTCTTGGGAATGCTTCCCTGACCTGGCCTTACACAGTG	6597
QY	4896	GGTCCTCTTCAAGTACTGTTTGTGAAGCTGGGCTCTTTGTGTAGTCCACCCACTGTA	5045
Db	65978	GGTCCTCTTCAAGTACTGTTTGTGAAGCTGGGCTCTTTGTGTAGTCCACCCACTGTA	6591
QY	5046	GGGCTACTGGGTTTAAAGGAACCTCCCCATTGGCAAAACCGGACCGGGCCGGCCAGG	5105
Db	65918	GGGCTACTGGGTTTAAAGGAACCTCCCCATTGGCAAAACCGGACCGGGCCGGCCAGG	6585
QY	5106	ACTGTGTTTCCAAAGGTTCCCGCCGCCCAACCCACAGCATAGCCGTGTAGTCCCTCGTG	5165
Db	65858	ACTGTGTTTCCAAAGGTTCCCGCCGCCCAACCCACAGCATAGCCGTGTAGTCCCTCGTG	6579
QY	5166	AGGCAGTGTGGTATATGTTCCACAGCACTGGGGGTCAAGACGCCCTTCTAGAACTTTCTAG	5225
Db	65798	AGGCAGTGTGGTATATGTTCCACAGCACTGGGGGTCAAGACGCCCTTCTAGAACTTTCTAG	6573
QY	5226	TTTGCCCTCTACCTGACCCCGACTTGTATTTCCCTTTTAGAGAGCCCTTCCCTCCGAGG	5285
Db	65738	TTTGCCCTCTACCTGACCCCGACTTGTATTTCCCTTTTAGAGAGCCCTTCCCTCCGAGG	6567
QY	5286	AGCCAAAGAGTGTGTGTGTGGCCATATTTGTGGCTGCTAATTTCACTGGTTCTTTTA	5345
Db	65678	AGCCAAAGAGTGTGTGTGTGGCCATATTTGTGGCTGCTAATTTCACTGGTTCTTTTA	6561
QY	5346	ATGTGAGGAACCTACATACTGACTTCAGTGGGACATCGGTGAGCGGGGCCGTGTGTGG	5405
Db	65618	ATGTGAGGAACCTACATACTGACTTCAGTGGGACATCGGTGAGCGGGGCCGTGTGTGG	6555
QY	5406	TGGGAGCCCTTTAGGGGACATGAGGACCTGGGGCCGCTGTGTGGTGGAGGCCAGGGCC	5465
Db	65558	TGGGAGCCCTTTAGGGGACATGAGGACCTGGGGCCGCTGTGTGGTGGAGGCCAGGGCC	6549
QY	5466	TCTCCCTTATAGTGGAGCAGAGTGTGTGGGCCCGCCGATGCTACATGCTGTGATTAAGAAAGG	5525
Db	65498	TCTCCCTTATAGTGGAGCAGAGTGTGTGGGCCCGCCGATGCTACATGCTGTGATTAAGAAAGG	6543
QY	5526	CTGAGTGTGTGACACCAAAACATGCGGACGAGGAGGCGCTGTGTGCCGCTTCCACA	5585
Db	65438	CTGAGTGTGTGACACCAAAACATGCGGACGAGGAGGCGCTGTGTGCCGCTTCCACA	6537
QY	5586	AGGACAGCCCTCTTACCCCTGAAAGGAACACTGGCTTGAAGGACTGAGACAGGCTGTG	5645
Db	65378	AGGACAGCCCTCTTACCCCTGAAAGGAACACTGGCTTGAAGGACTGAGACAGGCTGTG	6531
QY	5646	AGGAGCAGCCCTCTCTACAGGAGAGGACACAGTGTGGCCACAGTCTACTGCTGAGGTGC	5705
Db	65318	AGGAGCAGCCCTCTCTACAGGAGAGGACACAGTGTGGCCACAGTCTACTGCTGAGGTGC	6525
QY	5706	TTCTTCACACAGGAAACCGCGCAGCTGTGACTCGCTTGAGATGGGAAGCGGCGCCACA	5765

[illegible]

QY 3246 AAGGTGAGGGCTACATGCCCTCTCTGTACAGCAGTTCAGGGGATGCTGCTCTTC 3305  
|||||  
Db 67718 AAGGTGAGGGCTACATGCCCTCTCTGTACAGCAGTTCAGGGGATGCTGCTCTTC 67659  
QY 3306 AAGACAGCCCTGCCCTATGACAAAGACCAAGTCTTCAAGTTCAATGAGGACAGCCAGC 3365  
|||||  
Db 67658 AAGACAGCCCTGCCCTATGACAAAGACCAAGTCTTCAAGTTCAATGAGGACAGCCAGC 67599  
QY 3366 GTCTGGGAGAGAGATGAGAGAGAGGAGAAAGTGTCTCCCAAGGTTCCCTAGGCAATTCGA 3425  
|||||  
Db 67598 GTCTGGGAGAGAGATGAGAGAGAGGAGAAAGTGTCTCCCAAGGTTCCCTAGGCAATTCGA 67539  
QY 3426 GGAAGCTTGGGACATCTGCTGTGTGGTGGCCAGAGCCTCTGCTGGAAGGGGAGCAGAGA 3485  
|||||  
Db 67538 GGAAGCTTGGGACATCTGCTGTGTGGTGGCCAGAGCCTCTGCTGGAAGGGGAGCAGAGA 67479  
QY 3486 GGAAGTGAAGGAAACCCGCTGCCCTTATCTTGAAGTACGCCACACTGGGCTGGAGCCCTG 3545  
|||||  
Db 67478 GGAAGTGAAGGAAACCCGCTGCCCTTATCTTGAAGTACGCCACACTGGGCTGGAGCCCTG 67419  
QY 3546 GGGGAGATCCCGGGGGTTCGCCACAGGAGCACTGACTGATGACTTACACTGAGACTGT 3605  
|||||  
Db 67418 GGGGAGATCCCGGGGGTTCGCCACAGGAGCACTGACTGATGACTTACACTGAGACTGT 67359  
QY 3606 GGGGACTCTGACAGTCACTACACCGTTCTGACGCCAGAGCAGCTGGTGGTGGT 3665  
|||||  
Db 67358 GGGGACTCTGACAGTCACTACACCGTTCTGACGCCAGAGCAGCTGGTGGTGGT 67299  
QY 3666 TACATTCATTAACACTATATGATATTTAAAGAGAAAGTTTCAGATTTTGCATTCGA 3725  
|||||  
Db 67298 TACATTCATTAACACTATATGATATTTAAAGAGAAAGTTTCAGATTTTGCATTCGA 67239  
QY 3726 AGGCTTATTTATATATATGTGTGTATATATATATATATATATATATATATATATATAT 3785  
|||||  
Db 67238 AGGCTTATTTATATATATGTGTGTATATATATATATATATATATATATATATATATAT 67179  
QY 3786 ATTTTGGCTGGGAGAGTGTGAGTTTGGCTTCTTAAGGAGGAGCCGGCAGGCTCTT 3845  
|||||  
Db 67178 ATTTTGGCTGGGAGAGTGTGAGTTTGGCTTCTTAAGGAGGAGCCGGCAGGCTCTT 67119  
QY 3846 TGTCTGTATCTGTGGGAGAGATGGGTCTGGCTTGTGTCTGCTGCTTATCTTAAAGAT 3905  
|||||  
Db 67118 TGTCTGTATCTGTGGGAGAGATGGGTCTGGCTTGTGTCTGCTGCTTATCTTAAAGAT 67059  
QY 3906 CATCTCCATCTCTCCAGCGCATCTGTGTGACAGACCAACCAAGGAGATGAACCTGGCC 3965  
|||||  
Db 67058 CATCTCCATCTCTCCAGCGCATCTGTGTGACAGACCAACCAAGGAGATGAACCTGGCC 66999  
QY 3966 CTCTTGGGCGCTGACAGAGTCTCTTACCTTCTGTGTGCTGCTGCTGCTGCTGCTGCT 4025  
|||||  
Db 66998 CTCTTGGGCGCTGACAGAGTCTCTTACCTTCTGTGTGCTGCTGCTGCTGCTGCTGCT 66939  
QY 4026 AAGTCACATCTCTTCCACAGTAAATCCCTGGGAGCGCTGACCCCTGGGCTGCTGACG 4085  
|||||  
Db 66938 AAGTCACATCTCTTCCACAGTAAATCCCTGGGAGCGCTGACCCCTGGGCTGCTGACG 66879  
QY 4086 TTCTCTGTCTGGGCGCAGCATTTTGAAGATTTATCTTGAAGCGAGGCTGGCTCCCT 4145  
|||||  
Db 66878 TTCTCTGTCTGGGCGCAGCATTTTGAAGATTTATCTTGAAGCGAGGCTGGCTCCCT 66819  
QY 4146 ACTTATCCCTGCTCTCCCATTTCTCTGTGTGAGAGAGATGAGAGCAAGAGATGA 4205  
|||||  
Db 66818 ACTTATCCCTGCTCTCCCATTTCTCTGTGTGAGAGAGATGAGAGCAAGAGATGA 66759  
QY 4206 GAAAGATAGGGGCTGAAGAGCGACCTCCAGATGGCTCTTCTATCTGCTCTTCTGTT 4265  
|||||  
Db 66758 GAAAGATAGGGGCTGAAGAGCGCACCTCCAGATGGCTCTTCTATCTGCTCTTCTGTT 66699  
QY 4266 GAAACACAGCTGTGTGGGCTCAGGGCTTCTGAAGTGTCTTCTGATGATGAGACAG 4325  
|||||  
Db 66698 GAAACACAGCTGTGTGGGCTCAGGGCTTCTGAAGTGTCTTCTGATGATGAGACAG 66639

QY 4326 AGATCAGACAGCTGTACATCTGCTGTGTCTGAAGTGTGTTGCAAGCTTCACTCTCTCC 4385  
|||||  
Db 66638 AGATCAGACAGCTGTACATCTGCTGTGTCTGAAGTGTGTTGCAAGCTTCACTCTCTCC 66579  
QY 4386 CTAGTGTAGAGCAAGCCAGTGTCTTGTGAGAAACCAACCGGCGGGGAGATTTTAC 4445  
|||||  
Db 66578 CTAGTGTAGAGCAAGCCAGTGTCTTGTGAGAAACCAACCGGCGGGGAGATTTTAC 66519  
QY 4446 AGCAAGGCGCTCTCCCTTGGGATTAATCTTGTGAATTCACCTTCCCGGCTCTGTC 4505  
|||||  
Db 66518 AGCAAGGCGCTCTCCCTTGGGATTAATCTTGTGAATTCACCTTCCCGGCTCTGTC 66459  
QY 4506 TGGAGCCCATCTCTGTGTATCTGTGTGTTTGGACCCCTAATGTACGTTGGCTGTAG 4565  
|||||  
Db 66458 TGGAGCCCATCTCTGTGTATCTGTGTGTTTGGACCCCTAATGTACGTTGGCTGTAG 66399  
QY 4566 ACTCCCGAGAGTTGGTATGTGTGTAGAACAAATGGAGGCTGTGATGGTGTAGATTC 4625  
|||||  
Db 66398 ACTCCCGAGAGTTGGTATGTGTGTAGAACAAATGGAGGCTGTGATGGTGTAGATTC 66339  
QY 4626 ACATCCAGCCTTGGAAATCTAACCGGCATTACAAACCGAGTTACACCTTCCACTCCCTG 4685  
|||||  
Db 66338 ACATCCAGCCTTGGAAATCTAACCGGCATTACAAACCGAGTTACACCTTCCACTCCCTG 66279  
QY 4686 CTTAGGATCTGTCCCTGGGCTGGAACCTGAATAAGCTAATTTTGGGCTACGGTGGC 4745  
|||||  
Db 66278 CTTAGGATCTGTCCCTGGGCTGGAACCTGAATAAGCTAATTTTGGGCTACGGTGGC 66219  
QY 4746 AGTAGGGAACCTAGAGAGGTGTGAGTGGCATTTTGCAGAGATTTAGCCATGAGGTGT 4805  
|||||  
Db 66218 AGTAGGGAACCTAGAGAGGTGTGAGTGGCATTTTGCAGAGATTTAGCCATGAGGTGT 66159  
QY 4806 TCTTGAACCTTACTTCTGTGAAGTGTGAGTGTGACTGTGAAGTTTCTAGCACTGAACA 4865  
|||||  
Db 66158 TCTTGAACCTTACTTCTGTGAAGTGTGAGTGTGACTGTGAAGTTTCTAGCACTGAACA 66099  
QY 4866 AAGCTCAGGTTGTCTGCTGTCTATGACATGCTTAAAGCAGTCTGCTCTTCTGAGCTT 4925  
|||||  
Db 66098 AAGCTCAGGTTGTCTGCTGTCTATGACATGCTTAAAGCAGTCTGCTCTTCTGAGCTT 66039  
QY 4926 TGGCATCTGTGTCTTCTTGTGAATGAGTGTCTCTGACCTGCTGTACACGCTG 4985  
|||||  
Db 66038 TGGCATCTGTGTCTTCTTGTGAATGAGTGTCTCTGACCTGCTGTGTACACGCTG 65979  
QY 4986 GGTCTCTTCAAGTACTGTTTGAAGTGTGGCTCTTGTGTAGTCTCCACCACTGTGA 5045  
|||||  
Db 65978 GGTCTCTTCAAGTACTGTTTGAAGTGTGGCTCTTGTGTAGTCTCCACCACTGTGA 65919  
QY 5046 GGGCTAGCTGGGCTTAAGGGAACCTCTCCCATTTGGCAACCGGACCGGCGCGCAGG 5105  
|||||  
Db 65918 GGGCTAGCTGGGCTTAAGGGAACCTCTCCCATTTGGCAACCGGACCGGCGCGCAGG 65859  
QY 5106 ACTGTGTTTCCAAAGGTTCCCGCCCAACCCAGATCGGCTGATAGCTCCGCTCTGCT 5165  
|||||  
Db 65858 ACTGTGTTTCCAAAGGTTCCCGCCCAACCCAGATCGGCTGATAGCTCCGCTCTGCT 65799  
QY 5166 AGGCAAGTGTGTATGTTCAGAGAGTGGGGTGAAGCGCCCTTCTCAGAACTTTTAG 5225  
|||||  
Db 65798 AGGCAAGTGTGTATGTTCAGAGAGTGGGGTGAAGCGCCCTTCTCAGAACTTTTAG 65739  
QY 5226 TTGCCCTCTACCTGACCTGACCTGATTTCTTGAAGTGTGAGTGTGCTGCTGCTGCTGCT 5285  
|||||  
Db 65738 TTGCCCTCTACCTGACCTGACCTGATTTCTTGAAGTGTGAGTGTGCTGCTGCTGCTGCT 65679  
QY 5286 AGCCAAAGAGTGTGTGTGTGGGCTATATGTGTGAGCTGATTTCAATCTGTTCTTTTA 5345  
|||||  
Db 65678 AGCCAAAGAGTGTGTGTGTGGGCTATATGTGTGAGCTGATTTCAATCTGTTCTTTTA 65619  
QY 5346 ATGTGAGAACTACATACTGACTTCAAGTGGGAGCTGGTGGGCGGCTGTGTGTG 5405  
|||||  
Db 65618 ATGTGAGAACTACATACTGACTTCAAGTGGGAGCTGGTGGGCGGCTGTGTGTG 65559  
QY 5406 TGGGACCCCTTTAGCGGAGCTAGTGAAGTGGGCGGCTGTGTGTGTGAGCCAGGCGC 5465

Db	Accession	Sequence	Position
Db 65558	TGGGACCCCCCTTTACGGGACTCACTGAGCTGGGGCCCTCTGTGTGGTGAAGCCAGGGCC	65439	
QY 5466	TTCTCCCTTTAGTGGAGCCAGGTTGTCTGGGGCCCCGAATGTCACTGGTGAATCTAAGAAGG	5525	
Db 65498	TTCTCCCTTTAGTGGAGCCAGGTTGTCTGGGGCCCCGAATGTCACTGGTGAATCTAAGAAGG	65439	
QY 5526	CTGAGTGGTCTGACACCAAAACATGCCGACAGGAGGGGTGTGTGCCGCTGCTTCCACAA	5585	
Db 65438	CTGAGTGGTCTGACACCAAAACATGCCGACAGGAGGGGTGTGTGCCGCTGCTTCCACAA	65379	
QY 5586	AGGACAGCCCTCTCTTGACACCTTGAAAGAACTGGCTTTGAAGGACTGCAGACAGGCTGTG	5645	
Db 65378	AGGACAGCCCTCTCTTGACACCTTGAAAGAACTGGCTTTGAAGGACTGCAGACAGGCTGTG	65319	
QY 5646	AGGGGACAGCCCTCTCTGACGCGAGGACAGCAAGGTGGCCACAGTGTCACTGGTCAAGTGC	5705	
Db 65318	AGGGGACAGCCCTCTCTGACGCGAGGACAGCAAGGTGGCCACAGTGTCACTGGTCAAGTGC	65259	
QY 5706	TTTCTCACCACGGGAAAGCGCGGACCTGTGACTGCTTGTGAGTGGGAAAGCGGCCACAA	5765	
Db 65258	TTTCTCACCACGGGAAAGCGCGGACCTGTGACTGCTTGTGAGTGGGAAAGCGGCCACAA	65139	
QY 5766	GACCCCGGGTCTCTTGAGCTCTGTGTGGGGCCGCCCTTGCCACCTTTGTCTGTGGCTCAG	5825	
Db 65198	GACCCCGGGTCTCTTGAGCTCTGTGTGGGGCCGCCCTTGCCACCTTTGTCTGTGGCTCAG	65139	
QY 5826	GGTGCAGGAGAGCGCCCTCTCTGTGGTGGGCGGGCTTGTGTGCCGCTTGTGGGCTGTCTTA	5885	
Db 65138	GGTGCAGGAGAGCGCCCTCTCTGTGGTGGGCGGGCTTGTGTGCCGCTTGTGGGCTGTCTTA	65079	
QY 5886	CCATPAACCCGTTCCCGAGGGCTCTGACAGGCACTGTGACGGCTGGCTCCCTGGGCACTGCT	5945	
Db 65078	CCATPAACCCGTTCCCGAGGGCTCTGACAGGCACTGTGACGGCTGGCTCCCTGGGCACTGCT	65019	
QY 5946	CTCTCGGTGTGACACTGTGCTCAGGCGCAGGCGTCACACACTGGGGTCCGTCCGGAAGAT	6005	
Db 65018	CTCTCGGTGTGACACTGTGCTCAGGCGCAGGCGTCACACACTGGGGTCCGTCCGGAAGAT	64959	
QY 6006	GGGATCTTTCTGGGAGCTGCGCCCGGACAGAGTGGGAGCTCTTAATTTGTGGGGGAAGC	6065	
Db 64958	GGGATCTTTCTGGGAGCTGCGCGCGGACAGAGTGGGAGCTCTTAATTTGTGGGGGAAGC	64899	
QY 6066	TTTGATATCCATATGCGACGTCATTCACCCACCCCTTTTCTGTACAGACACATAGTCTT	6125	
Db 64898	TTTGATATCCATATGCGACGTCATTCACCCACCCCTTTTCTGTGTGTGTGTGTGTGTGT	64839	
64838	ACATTGGATTTTGTAAATAAATAAATAAATAAATGAAGACTTTAACTC 64792		
RESULT 9	HSU96629/c	167343 bp DNA linear PRI 10-JAN-2000	
LOCUS	Human chromosome 8 BAC clone C19875K-2A8 complete sequence.		
DEFINITION	U96629		
VERSION	U96629.1 GI:2341008		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Loftus, B.-J., Kim, U.-J., Sneddon, V. P., Kalush, F., Brandon, R., Fulmerman, J., Mason, T., Crosby, M. L., Barnstead, M., Cronin, L., Deslauriers, Mays, A., Cao, Y., Xu, R. X., Kang, H. L., Mitchell, S., Eichler, E. E., Harris, P. C., Venter, J. C. and Adams, M. D.		
TITLE	Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q		
JOURNAL	Genomics 60 (3), 295-308 (1999)		
MEDLINE	99425270		
PUBMED	10493829		

REFERENCE	TITLE
2 (bases 1 to 167343)	Adams,M.D.
HUMAN CHROMOSOME 8 BAC CLONE C1987SK-2A8 COMPLETE SEQUENCE	Unpublished
3 (bases 1 to 167343)	Adams,M.D.
DIRECT SUBMISSION	Submitted (07-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 167343)	Adams,M.D., Loftus,B.J., Zhou,L., La Bombard,M., Kim,U.J. and Venter,J.C.
DIRECT SUBMISSION	Submitted (22-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
On Aug 22, 1997 this sequence version replaced gi:1930148.	BAC CLONE C1987SK-2A8 is located on chromosome 8. Genes were identified by a combination of five methods: XGRAIL (available by anonymous ftp from arthur.epm.gov), GeneFinder (available by anonymous ftp from collnuc.washington.edu), GENSCAN (available e-mail server at genescan.genome.stanford.edu), searches of the EST database at TIGR ( <a href="http://www.tigr.org/tdb/hcd/hcd.html">http://www.tigr.org/tdb/hcd/hcd.html</a> ) and searches against a peptide database. Repeats were identified using Censor (Jurka, J., Klonowski, P., Dagman, V., Pelton, P. Censor-a program for the identification and elimination of repetitive elements from DNA sequences. Computers Chem 20: 119-121 (1996); available by anonymous ftp from ncbi.nlm.nih.gov).
location/Qualifiers	1. 167343
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="8"	
/clone="2A8"	
complement(20..308)	/rpt_family="Alu-Jo"
complement(370..1051)	/rpt_family="L1"
1070..1355	/rpt_family="Alu-Jo"
complement(1382..2775)	/rpt_family="L1"
complement(3086..3383)	/rpt_family="Alu-Jo"
3861..3991	/rpt_family="Alu-Jo"
complement(4198..4490)	/rpt_family="Alu-Jb"
complement(4878..5157)	/rpt_family="Alu-Sz"
complement(5650..5945)	/rpt_family="Alu-Sx"
complement(7305..7586)	/rpt_family="Alu-Sz"
complement(7613..7717)	/rpt_family="L1MB3"
8787..9080	/rpt_family="Alu-Jo"
complement(9234..9495)	/rpt_family="L1MB3"
complement(9630..9902)	/rpt_family="Alu-Y"
complement(9914..10017)	/rpt_family="L1MB7"
10304..10598	/rpt_family="Alu-Sq"
11339..11669	/rpt_family="Alu-Jb"
complement(12169..12457)	/rpt_family="Alu-Sz"
complement(12844..13048)	/rpt_family="L1MB7"
complement(13100..13188)	/rpt_family="Alu-Jb"

repeat_region	/rpt_family="Alu-Jo"	28831..29113
repeat_region	/rpt_family="Alu-Sx"	complement(29159..29615)
repeat_region	/rpt_family="L1MA9"	complement(29616..29892)
repeat_region	/rpt_family="Alu-Sg"	complement(29911..29941)
repeat_region	/rpt_family="L1MA9"	29965..30243
repeat_region	/rpt_family="Alu-Sx"	complement(30252..30477)
repeat_region	/rpt_family="L1MA10"	30492..30781
repeat_region	/rpt_family="Alu-Sp"	complement(30782..31066)
repeat_region	/rpt_family="L1MA10"	complement(31067..31170)
repeat_region	/rpt_family="L1"	complement(31174..31463)
repeat_region	/rpt_family="Alu-Sz"	complement(31464..32119)
repeat_region	/rpt_family="L1"	complement(32156..32238)
repeat_region	/rpt_family="L1MA2"	32253..32537
repeat_region	/rpt_family="Alu-Y"	complement(32539..33095)
repeat_region	/rpt_family="L1MA2"	33193..33469
repeat_region	/rpt_family="Alu-Jo"	complement(33488..33809)
repeat_region	/rpt_family="MER42C"	complement(33812..34098)
repeat_region	/rpt_family="Alu-Sg"	
Query Match		
Best Local Similarity 98.6%; Score 3035; DB 9; Length 167343;		
Matches 3062; Conservative 0; Mismatches 45; Indels 0; Gaps 0;		
QY 3066	GTGGATGATATACATCACTGAGACACTTGGCCATGAACTTCCTGTCCTCCACATCACT	3125
Db 67898	GTGTGGATGCGCTCAGTCGCAAGGTTCTGTACGTGGATTCCTCCACTCTGTC	67839
QY 3126	CGAAGCCCCCATCAAGTGCACGTGACGCGTGGACATTCGATGCCAGATGCCCTCAG	3185
Db 67838	TCTCTCTCCCGTTCCAGAGTACCTCAGCGTGGACATTCGATGCCAGATGCCCTCAG	67779
QY 3186	GCCGTGTCATGATGACTCCACTTCACAGCGGCGCAAGTGCATCACTTCTTGTC	3245
Db 67778	GCCGTGTCATGATGACTCCACTTCACAGCGGCGCAAGTGCATCACTTCTTGTC	67719
QY 3246	AAGGTGACGGCTACATGCCCCCTCTGTACACGCAGTTCAAGGGTGTCTGTGCTTTC	3305
Db 67718	AAGGTGACGGCTACATGCCCCCTCTGTACACGCAGTTCAAGGGTGTCTGTGCTTTC	67659
QY 3306	AAGACACCCCTGCCCATGACACAGCAAGGCTTCAAGTTCATGAGGGGACGCGCAG	3365
Db 67658	AAGACACCCCTGCCCATGACACAGCAAGGCTTCAAGTTCATGAGGGGACGCGCAG	67599
QY 3366	GTCGTGGGAGAGGATGACGAGAGGAGAAATGGTCCCAAGTTCTTAGCATTTGCA	3425
Db 67598	GTCGTGGGAGAGGATGACGAGAGGAGAAATGGTCCCAAGTTCTTAGCATTTGCA	67539
QY 3426	GGACCTTGGGACATCTGCTGTGTGGTGCCCAAGGCTTCTGCTGGAGGGGACAGAGA	3485
Db 67538	GGACCTTGGGACATCTGCTGTGTGGTGCCCAAGGCTTCTGCTGGAGGGGACAGAGA	67479
QY 3486	GGAGTGAAGCAACCGCTTCCTTAACTTGAAGTCAAGCACCACTGGGCTGGAGCCTTG	3545
Db 67478	GGAGTGAAGCAACCGCTTCCTTAACTTGAAGTCAAGCACCACTGGGCTGGAGCCTTG	67419
QY 3546	GGCGAGATCCCGGGGTTTCCCAACAGGGGCACTGACTAATAGCTTACCTAGAGACTGT	3605

|||||  
Db 67418 GGGGGAGTCCCCGGGGTTCACACAGGGCAGTGTATACCTTAACATGTAGAGACTGT 67359  
QY 3606 GGGGACCTGACAGTACACACACGGCTTGAGCCAGACAGCGGTGCGGGTTT 3665  
Db 67358 GGGGACCTGACAGTACACACGGCTTGAGCCAGACAGCGGTGCGGGTTT 67299  
QY 3666 TACATTCATTAACACATATATATATTTAAAGAGAAAGTTTCAGATTTGCCATTCA 3725  
Db 67298 TACATTCATTAACACATATATATTTAAAGAGAAAGTTTCAGATTTGCCATTCA 67239  
QY 3726 AGCCTATTTATATATATGTTGTTGATATATAATACATGCACACACTTGCATATATAT 3785  
Db 67238 AGCCTATTTATATATATGTTGTTGATATATAATACATGCACACACTTGCATATATAT 67179  
QY 3786 ATTTTGGGCTGGGGAGTGTGAGTTTGGCTTCTAAGGAGGAGCCGACAGCTCTT 3845  
Db 67178 ATTTTGGGCTGGGGAGTGTGAGTTTGGCTTCTAAGGAGGAGCCGACAGCTCTT 67119  
QY 3846 TGTTCGTATTTCTGGCGAGATGGTCCCTGTGTCTGCTAGCTGCTTATCTTTAAAGAT 3905  
Db 67118 TGTTCGTATTTCTGGCGAGATGGTCCCTGTGTCTGCTAGCTGCTTATCTTTAAAGAT 67059  
QY 3906 CATCTCCCATCTCCCGACGGCCCATCTGTGTGCAGCAACCAAGAAAGGATGAAGTGGCC 3965  
Db 67058 CATCTCCCATCTCCCGACGGCCCATCTGTGTGCAGCAACCAAGAAAGGATGAAGTGGCC 66999  
QY 3966 CTCTTGGGGCTGGACAAAGTCTCTTACCTTCTGTTGCCAGTAGCAACCTGT 4025  
Db 66998 CTCTTGGGGCTGGACAAAGTCTCTTACCTTCTGTTGCCAGTAGCAACCTGT 66939  
QY 4026 AACTCACAATTCCTTCCCACTGAATCCCTGGAGCGCTGACCCCTGTGGGCTGTTCAGC 4085  
Db 66938 AACTCACAATTCCTTCCCACTGAATCCCTGGAGCGCTGACCCCTGTGGGCTGTTCAGC 66879  
QY 4086 TTCTCTGCTGCTGGGGCCAGGATTTTGGAGATTTATCTTAAAGCCAGCTGCTCCGT 4145  
Db 66878 TTCTCTGCTGCTGGGGCCAGGATTTTGGAGATTTATCTTAAAGCCAGCTGCTCCGT 66819  
QY 4146 ACTTATCCCTGCTCTCCCATTTTCTCTTGTGGAGAGAAATGAGAAAGCAAGATGTA 4205  
Db 66818 ACTTATCCCTGCTCTCCCATTTTCTCTTGTGGAGAGAAATGAGAAAGCAAGATGTA 66759  
QY 4206 GAAAGAAATAGGGGCTGAAGACGCCACTCCAGATGCTCTTCTATCTGCTCTTCTGT 4265  
Db 66758 GAAAGAAATAGGGGCTGAAGACGCCACTCCAGATGCTCTTCTATCTGCTCTTCTGT 66699  
QY 4266 GAAACACACGTGCTGGGGCTCAAGGCGTTTCTAAGTGCCTTCTTGTGATTTGGACAGG 4325  
Db 66698 GAAACACACGTGCTGGGGCTCAAGGCGTTTCTAAGTGCCTTCTTGTGATTTGGACAGG 66639  
QY 4326 AGATCAGCAGCGTGCACATCTGCTGTGCTGAAGTGGTTTGCAGGCTCCTCTCC 4385  
Db 66638 AGATCAGCAGCGTGCACATCTGCTGTGCTGAAGTGGTTTGCAGGCTCCTCTCC 66579  
QY 4386 CTATGTTAGAGCAAGCCAGTGTCTTCTGAGAGAACCAACCCGGCTGGCCGGGAAAGTTTAC 4445  
Db 66578 CTATGTTAGAGCAAGCCAGTGTCTTCTGAGAGAACCAACCCGGCTGGCCGGGAAAGTTTAC 66519  
QY 4446 AGCAAGGCGCTGCTGGGATATTTCTTGTGTAAATTCACCTTCCCCCGCTCTGTGC 4505  
Db 66518 AGCAAGGCGCTGCTGGGATATTTCTTGTGTAAATTCACCTTCCCCCGCTCTGTGC 66459  
QY 4506 TGGAGCCCATCTCTGTGTTATCTGTGGTTTTTGGACCCCTTAATTCAGCTGGCTGAGG 4565  
Db 66458 TGGAGCCCATCTCTGTGTTATCTGTGGTTTTTGGACCCCTTAATTCAGCTGGCTGAGG 66399  
QY 4566 ACTCCCGAGGTTTGTATGTGCTGAACAATGGAGGCTGTGATTTGCTGTGAAGCTC 4625  
Db 66398 ACTCCCGAGGTTTGTATGTGCTGAACAATGGAGGCTGTGATTTGCTGTGAAGCTC 66339  
QY 4626 ACATCCAGCTTGGAAATCTAACGGGCAATTCACAACCCGAGTTACCACCTTTCACCTCCCTG 4685  
|||||

Db 66338 ACATCCAGCCTTGGAATCTAACGGGCAATTCACAACCCGAGTTTACCACCTTCCCTCCCTG 66229  
QY 4686 CTTAGATTCCTGTTCCCTGGGCTGAACACTGAATAAGCTAATTTTTTGGCTCAGGGTGC 4745  
Db 66278 CTTAGATTCCTGTTCCCTGGGCTGAACACTGAATAAGCTAATTTTTTGGCTCAGGGTGC 66219  
QY 4746 AGTAGGGGAACCTAGAGAGGCTGTGAGTGCATTTGTCAAGGATTTAGCCATGACGTGT 4805  
Db 66218 AGTAGGGGAACCTAGAGAGGCTGTGAGTGCATTTGTCAAGGATTTAGCCATGACGTGT 66159  
QY 4806 TCTTGAACCCCTACTTCTGTGAAGTGGAGTGTGACCTGGAAGTTTTCAGCACTGAACAA 4865  
Db 66158 TCTTGAACCCCTACTTCTGTGAAGTGGAGTGTGACCTGGAAGTTTTCAGCACTGAACAA 66099  
QY 4866 AAGCTCAGTTTGTCTGTCATGCATGCAATGCCCTTAAGCCAGTTCCGTTCCCTAGACCT 4925  
Db 66098 AAGCTCAGTTTGTCTGTCATGCATGCAATGCCCTTAAGCCAGTTCCGTTCCCTAGACCT 66039  
QY 4926 TGGCATCTGTGCTTCTATTTTCTTGGAAATACGTTCTCTGTACCTGCTGTACACGTG 4985  
Db 66038 TGGCATCTGTGCTTCTATTTTCTTGGAAATACGTTCTCTGTACCTGCTGTACACGTG 65979  
QY 4986 GGTCCCTTCAAGTACTGTTTGAAGCTGGGCTTTTGTGTAGCTCCACCCACCTGTA 5045  
Db 65978 GGTCCCTTCAAGTACTGTTTGAAGCTGGGCTTTTGTGTAGCTCCACCCACCTGTA 65919  
QY 5046 GGGCTAGCTGGGCTTAAGGGAACTCTCCCATTTGGCAAAACCGAGCCGGCCGCCAGG 5105  
Db 65918 GGGCTAGCTGGGCTTAAGGGAACTCTCCCATTTGGCAAAACCGAGCCGGCCGCCAGG 65859  
QY 5106 ACTGTGTTCCAAAGGTTCCCGCCGCCCAACCCAGATAGCCTGTAGCTCCCTGTCTG 5165  
Db 65858 ACTGTGTTCCAAAGGTTCCCGCCGCCCAACCCAGATAGCCTGTAGCTCCCTGTCTG 65799  
QY 5166 AGGCAAGTGGTATATGTTCCAGAGTGGGGGTGACAGCCGCTTCTCAGAACTTTTAG 5225  
Db 65798 AGGCAAGTGGTATATGTTCCAGAGTGGGGGTGACAGCCGCTTCTCAGAACTTTTAG 65739  
QY 5226 TTGCCCTCTACGTGACTCTGACTTGTATTCCTTTTGAAGAGTACCTTCTCCCTCGGG 5285  
Db 65738 TTGCCCTCTACGTGACTCTGACTTGTATTCCTTTTGAAGAGTACCTTCTCCCTCGGG 65679  
QY 5286 AGCCAAAGATGTGTGTGTGGGCTATATTGTGGCTGATTTTCATCTGTTCTTTTA 5345  
Db 65678 AGCCAAAGATGTGTGTGTGGGCTATATTGTGGCTGATTTTCATCTGTTCTTTTA 65619  
QY 5346 ATGTAGGAACCTCACAATCTGACTTCACTGAGTGTGAGTGTGAGCCGGGCGCTGTGTGG 5405  
Db 65618 ATGTAGGAACCTCACAATCTGACTTCACTGAGTGTGAGTGTGAGCCGGGCGCTGTGTGG 65559  
QY 5406 TGGGACCCCTTTTAGCGGAACTAGTGAAGCTGGGGCGCTGTGTGTGGAGCCAGGGCC 5465  
Db 65558 TGGGACCCCTTTTAGCGGAACTAGTGAAGCTGGGGCGCTGTGTGTGGAGCCAGGGCC 65499  
QY 5466 TTCTCCCTTTAGTGAAGCAGGTTGTGGGCCCCGAATGTCACTGTGTGATTAAGAAGG 5525  
Db 65498 TTCTCCCTTTAGTGAAGCAGGTTGTGGGCCCCGAATGTCACTGTGTGATTAAGAAGG 65439  
QY 5526 CTGAGTGTCTGACACCAAAACATGCCGAGGAGGAGGCTGTGGGCGGCTTCCAAACA 5585  
Db 65438 CTGAGTGTCTGACACCAAAACATGCCGAGGAGGAGGCTGTGGGCGGCTTCCAAACA 65379  
QY 5586 AGGACAGCCCTCTTGAACCCCTGAAGAGAACCTGGCTTGAAGACTGCAGACAGGCTGTG 5645  
Db 65378 AGGACAGCCCTCTTGAACCCCTGAAGAGAACCTGGCTTGAAGACTGCAGACAGGCTGTG 65319  
QY 5646 AGGGGCAAGCCCTCTCAGAGAGAGGAGCAAGGTGGCCACAGTGTCACTGTAGGTGC 5705  
Db 65318 AGGGGCAAGCCCTCTCAGAGAGAGGAGCAAGGTGGCCACAGTGTCACTGTAGGTGC 65259  
QY 5706 TTCTCAGCAGGGGAAGCCGGCCGACGTGTGACTGCTTGTGAGATGGGAAGGCGGCCACA 5765  
Db 65258 TTCTCAGCAGGGGAAGCCGGCCGACGTGTGACTGCTTGTGAGATGGGAAGGCGGCCACA 65199  
|||||



```

QY 5766 GACCCCGGCTCTCTGGCTGTGTGGGCGCCCGCCACCTTGTCTGGCTGCAG 5825
      |||
Db 65198 GACCCCGGCTCTCTGGCTGTGTGGGCGCCCGCCACCTTGTCTGGCTGCAG 65139
      |||
QY 5826 GGGGAGGAGGCGCTGTCTGTGGTGGCGGCTTGTCTGGCTGTCTGA 5885
      |||
Db 65138 GGGGAGGAGGCGCTGTCTGTGGTGGCGGCTTGTCTGGCTGTCTGA 65079
      |||
QY 5886 CCATAACACCGTCCAGGCGCTGCAGGCGCAGTGTGGCTGGCTGGCTGGCT 5545
      |||
Db 65078 CCATAACACCGTCCAGGCGCTGCAGGCGCAGTGTGGCTGGCTGGCTGGCT 65019
      |||
QY 5946 CCTCCGTGTGAGCTGTGCTCAGGCGGCTCACCAGCTGGCTGTCCGGAAGAT 6005
      |||
Db 65018 CCTCCGTGTGAGCTGTGCTCAGGCGGCTCACCAGCTGGCTGTCCGGAAGAT 64959
      |||
QY 6006 GGGATCTTCTGGAGCTGGCGGCGGAGACAGATGGGAGCTCTTGTGGGGGAGAC 6065
      |||
Db 64958 GGGATCTTCTGGAGCTGGCGGCGGAGACAGATGGGAGCTCTTGTGGGGGAGAC 64899
      |||
QY 6066 TTGTATATCCATCCACGCTCATCCACCCCTTTTGTGACAGACATATGCTCT 6125
      |||
Db 64898 TTGTATATCCATCCACGCTCATCCACCCCTTTTGTGACAGACATATGCTCT 64839
      |||
QY 6126 ACATGTGATTTTGTAAAAAATAAATAATGAAGACTTTTACTC 6172
      |||
Db 64838 ACATGTGATTTTGTAAAAAATAAATAATGAAGACTTTTACTC 64792
      |||

RESULT 10
AC040975/c 168194 bp DNA linear HTG 06-FEB-2002
LOCUS DEFINITION Homo sapiens chromosome 8 clone RP11-662B19 map 8, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.
AC040975
AC040975.6 GI:18543125
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE Birren, B., Linton, L., Nusbaum, C. and Landier, E.
JOURNAL Homo sapiens chromosome 8, clone RP11-662B19
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 168194)
      Birren, B., Linton, L., Nusbaum, C., Landier, E., Abraham, H., Allen, N.,
      Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
      Boguslavskiy, L., Bouknight, B., Brown, A., Burkett, G.,
      Campiolo, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
      Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
      Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
      Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
      Grand-Pierre, N., Grant, G., Hages, B., Heald, A., Horton, L.,
      Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
      Klein, J., Lakocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
      Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
      McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McSheeters, R.,
      Melarini, J., Meneses, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J.,
      Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
      O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pletre, N.,
      Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
      Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
      Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
      Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
      Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
      Young, G., Zainoun, J., Zimmer, A. and Zody, M.
      Direct Submission
      Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
      Research, 320 Charles Street, Cambridge, MA 02141, USA
      On Feb 6, 2002 this sequence version replaced gi:16554398.
      All repeats were identified using RepeatMasker:

```

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9374
Center clone name: 662_B_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 124099: contig of 124099 bp in length
*
124100 124199: gap of 100 bp
*
124200 166180: contig of 41981 bp in length
*
166181 166280: gap of 100 bp
*
166281 168194: contig of 1914 bp in length.
*
Location/Qualifiers
1. 168194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-662B19"
/clone_lib="RPCT-11 Human Male BAC"
BASE COUNT 48612 a 35313 c 35554 g 48474 t 241 others
ORIGIN

Query Match 48.9% Score 3018.4; DB 2; Length 168194;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 3059; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY 3066 GTGATGATATCATCACTGTGAGACATGCCATGATCTCTTGTCTCCACATCACT 3125
      |||
Db 147185 GTGTGGATGGCGCTCAGCTGCAAGGGTCTGTGATTTACTGTGGATTTCCACTCTGTG 147126
      |||
QY 3126 CGGAAGCCCCCATCAAGGTACTCAGGTGACATTCGATGCCAGATGCCCTCAG 3185
      |||
Db 147125 TCTCTCCCGCTTCCAGGTGACCTCAAGGTGACATTCGATGCCAGATGCCCTCAG 147066
      |||
QY 3186 GCCCTGTCTATGATGACTCCCACTTCCAGAGGGGAGCAAGTGCATCACTTCTTCTG 3245
      |||
Db 147065 GCCCTGTCTATGATGACTCCCACTTCCAGAGGGGAGCAAGTGCATCACTTCTTCTG 147006
      |||
QY 3246 AAGGTGAGGCTACATGACCCCTCTGTACAGCAGTTCAGGGTGTGCTCTTCTG 3305
      |||
Db 147005 AAGGTGAGGCTACATGACCCCTCTGTACAGCAGTTCAGGGTGTGCTCTTCTG 146946
      |||
QY 3306 AAGACACGCTTGGCCCATGCAAGACCAAGTCTTCAAGTTCACTAAGGGGAGCGCAGC 3365
      |||
Db 146945 AAGACACGCTTGGCCCATGCAAGACCAAGTCTTCAAGTTCACTAAGGGGAGCGCAGC 146886
      |||
QY 3366 GTCTGGGAGAGAGATGAGCAGAGGGGAGAAATGGCTCCCAAGTCTTCAAGTTCGA 3425
      |||
Db 146885 GTCTGGGAGAGAGATGAGCAGAGGGGAGAAATGGCTCCCAAGTCTTCAAGTTCGA 146826
      |||
QY 3426 GGAACCTTGGGACATCTGTGTGTGTGGCCCGACGCTCTGTGGAAGGGGAGCAGAGA 3485
      |||
Db 146825 GGAACCTTGGGACATCTGTGTGTGTGGCCCGACGCTCTGTGGAAGGGGAGCAGAGA 146766
      |||
QY 3486 GGAAGTGAAGAAACCGCTGCTTATCTTGAAGTCAAGCCACACTGGGCTGGAGCCCTG 3545
      |||
Db 146765 GGAAGTGAAGAAACCGCTGCTTATCTTGAAGTCAAGCCACACTGGGCTGGAGCCCTG 146706
      |||
QY 3546 GCGGAGTCCCGGGGGTTCCTCCACACAGGCGCACTGATGATGACTTACACTGAGACTGT 3605
      |||

```





Dh 197884 GCTGCTCCAAACCCGATCCGCTCAGCGTCAAGCTTCACGCTCTTTGTATCTGTCT 197943  
Qy 718 TCTTCCGCTCATCGCCCACTATTACCTCACACTGTGATGAGGCTGATGAGCGACGA 777  
Dh 197944 TCTTCCGCTCATCGCCCACTATTACCTCACACTGTGATGAGGCTGATGAGCGACGA 198003  
Qy 778 AGCGGATTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAACGACGCTGTGATC 837  
Dh 198004 AGCGGATTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAACGACGCTGTGATC 198063  
Qy 838 TGTGCGCATCCGGGAGTGGGTGATGAAGAGGCTCCGACGTGGAGGCCAAGCGCCAG 897  
Dh 198064 TGTGCGCATCCGGGAGTGGGTGATGAAGAGCTCCGACGTGGAGGCCAAGCGCCAG 198123  
Qy 898 AGCTGAACAGCAGAGATCGCCAGCTGATCTGAAGATCGAAGCTGTGAAGAGCATTTG 957  
Dh 198124 AGCTGAACAGCAGAGATCGCCAGCTGATCTGAAGATCGAAGCTGTGAAGAGCATTTG 198183  
Qy 958 AGAAGCGCAAGCAGAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGGACATTCCT 1017  
Dh 198184 AGAAGCGCAAGCAGAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGGACATTCCT 198243  
Qy 1018 ACAAGAGCTCATGCGCCAGACAGCCAGCGTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 1077  
Dh 198244 ACAAGAGCTCATGCGCCAGACAGCCAGCGTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 198303  
Qy 1078 AGAAGAGCTATGCGCGGCTCCCTCCGCGAAGGCACTCGGGGCTGCGGCTACACAACT 1137  
Dh 198304 AGAAGAGCTATGCGCGGCTCCCTCCGCGAAGGCACTCGGGGCTGCGGCTACACAACT 198363  
Qy 1138 GCTTTGATTTATTTCTCGTGGCTCTCAGCTCTGCGCTTCCGGTCTACGCTTATGACAGTG 1197  
Dh 198364 GCTTTGATTTATTTCTCGTGGCTCTCAGCTCTGCGCTTCCGGTCTACGCTTATGACAGTG 198423  
Qy 1198 ACCAGTTTCTTTGGAGCTACCTGGATCCCTGGTCAAGCAAGGCTTTTCAAGCGCAG 1257  
Dh 198424 ACCAGTTTCTTTGGAGCTACCTGGATCCCTGGTCAAGCAAGGCTTTTCAAGCGCAG 198483  
Qy 1258 CAGAGCTAACGTTTATGTTACAGAAATGACAGATCGGCTGCTTTTACGTATAGTAG 1317  
Dh 198484 CAGAGCTAACGTTTATGTTACAGAAATGACAGATCGGCTGCTTTTACGTATAGTAG 198543  
Qy 1318 TGGGAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGCTGAGAGACAGTGTATT 1377  
Dh 198544 TGGGAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGCTGAGAGACAGTGTATT 198603  
Qy 1378 CCGTGGCAGCTGCGGAGGAGTGAACACAACTGTCATCATCATCTGTCAAGTAAGT 1437  
Dh 198604 CCGTGGCAGCTGCGGAGGAGTGAACACAACTGTCATCATCATCTGTCAAGTAAGT 198663  
Qy 1438 CAGATACACAGAACTTCTATAGCTACGTACGTGCGCTGCTGATGTCGCGCATGCA 1497  
Dh 198664 CAGATACACAGAACTTCTATAGCTACGTACGTGCGCTGCTGATGTCGCGCATGCA 198723  
Qy 1498 CATTACACAGTGTGAGTACAGAGCTGCTTGAATGCTGCTGATACAGCCGCTGCTCATG 1557  
Dh 198724 CATTACACAGTGTGAGTACAGAGCTGCTTGAATGCTGCTGATACAGCCGCTGCTCATG 198783  
Qy 1558 CCAATGCTGAGCCCAATTCTCATGGAATCCACACAGAGTGGCGGTGAACCGGAAATATC 1617  
Dh 198784 CCAATGCTGAGCCCAATTCTCATGGAATCCACACAGAGTGGCGGTGAACCGGAAATATC 198843  
Qy 1618 TCTTCACTTTCAGAGGCGGAGAGATGAGTCTGTAGAGTCTAGCCCTTTCAGAGGCGCGCT 1677  
Dh 198844 TCTTCACTTTCAGAGGCGGAGAGATGAGTCTGTAGAGTCTAGCCCTTTCAGAGGCGCGCT 198903  
Qy 1678 CATTGCAAGGAGAAATGAGGAGGCGCTCCCGGCTGCTGATGAGAGGAGTATGCGCA 1737  
Dh 198904 CATTGCAAGGAGAAATGAGGAGGCGCTCCCGGCTGCTGATGAGAGGAGTATGCGCA 198963  
Qy 1738 CCGTGAAGGCGGTGCAAGAGCAGCAAGCTGATAGCTCTGCTGGAATTAACCTGCAAAA 1797  
Dh 198964 CCGTGAAGGCGGTGCAAGAGCAGCAAGCTGATAGCTCTGCTGGAATTAACCTGCAAAA 199023

Qy 1798 ACCAGCCCAAAACCAGCTGCGAGCTGAGTGGGCACTGTGTGGAGAGCGGAGACCGCT 1857  
Dh 199024 ACCAGCCCAAAACCAGCTGCGAGCTGAGTGGGCACTGTGTGGAGAGCGGAGACCGCT 199083  
Qy 1858 TGAATTTGCTGAAGCTCTCCACTTGGCCCTCATTTATTTACCCCCGGGAGACCTCGCTTGG 1917  
Dh 199084 TGAATTTGCTGAAGCTCTCCACTTGGCCCTCATTTATTTACCCCCGGGAGACCTCGCTTGG 199143  
Qy 1918 TTAATTTCTGTGGGTGTGCAACAGGCTCTTGAACCCCTGGAAAGTCCGCTGCCG 1977  
Dh 199144 TTAATTTCTGTGGGTGTGCAACAGGCTCTTGAACCCCTGGAAAGTCCGCTGCCG 199203  
Qy 1978 TGGTGTGGGGGAGCAGGTCCAGCTTCCCTACAGAGATGCTGCAAGTGAACGAGGCG 2037  
Dh 199204 TGGTGTGGGGGAGCAGGTCCAGCTTCCCTACAGAGATGCTGCAAGTGAACGAGGCG 199263  
Qy 2038 CCGTGGTGGTCCAAAGCCTCGTGTACCGAGGTTATTTCTGCTCAGAGCCTCTCG 2097  
Dh 199264 CCGTGGTGGTCCAAAGCCTCGTGTACCGAGGTTATTTCTGCTCAGAGCCTCTCG 199323  
Qy 2098 ATAGTACCTCTGCGCTATGAGGCGGCAAGGCGCTTCTGCGGAGCTTACTTCTCA 2157  
Dh 199324 ATAGTACCTCTGCGCTATGAGGCGGCAAGGCGCTTCTGCGGAGCTTACTTCTCA 199383  
Qy 2158 CTGCTACAGATATTTTAAATACCGTCTGCTATGATTTAGAGTCCGATCCAGATCCAG 2217  
Dh 199384 CTGCTACAGATATTTTAAATACCGTCTGCTATGATTTAGAGTCCGATCCAGATCCAG 199443  
Qy 2218 CCGCTCCCATCCGGGAAGGCGGCAAGCTGATGATCCCAACGCTTACAGCAGGCGCTG 2277  
Dh 199444 CCGCTCCCATCCGGGAAGGCGGCAAGCTGATGATCCCAACGCTTACAGCAGGCGCTG 199503  
Qy 2278 GAACGTACCCCAACATGCGTGAACAGGAGGAGCTGGAACGCGGAGGAGGAGAGC 2337  
Dh 199504 GAACGTACCCCAACATGCGTGAACAGGAGGAGCTGGAACGCGGAGGAGGAGAGC 199563  
Qy 2338 CGCCCTACGCTTACCCAGATACCTCCGCAATTTCACTGTGACTGTACGTATTTAAC 2397  
Dh 199564 CGCCCTACGCTTACCCAGATACCTCCGCAATTTCACTGTGACTGTACGTATTTAAC 199623  
Qy 2398 GCAGCTGGAAGTGTGCTCCAGAGGCTTTTCATTTTCCCCCACTCCCTTTGACCCG 2457  
Dh 199624 GCAGCTGGAAGTGTGCTCCAGAGGCTTTTCATTTTCCCCCACTCCCTTTGACCCG 199683  
Qy 2458 TGTGCTCCCTAGAGGCAATTTTGGGCTCAGAGACTGGCTTTCGCGCTATTTGAGTG 2517  
Dh 199684 TGTGCTCCCTAGAGGCAATTTTGGGCTCAGAGACTGGCTTTCGCGCTATTTGAGTG 199743  
Qy 2518 GAGCTGGGGTCTGCGAAGAAATTTGAGGAGCGCTTGGAGGCAATGTTCCCGAGAGC 2577  
Dh 199744 GAGCTGGGGTCTGCGAAGAAATTTGAGGAGCGCTTGGAGGCAATGTTCCCGAGAGC 199803  
Qy 2578 AGTTCAAGGAGTGTGATGATGATGATGAGGAGGAGGAGGCTTATGAACTGTTTGA 2637  
Dh 199804 AGTTCAAGGAGTGTGATGATGATGATGAGGAGGAGGAGGCTTATGAACTGTTTGA 199863  
Qy 2638 GGCTGATGAGCTTCCCTTACCTGAACAGGCTGCTGCTGCTGAGAAATTC 2686  
Dh 199864 GGCTGATGAGCTTCCCTTACCTGAACAGGCTGCTGCTGCTGAGAAATTC 199912

RESULT 12  
AF083550 3479 bp mRNA linear ROD 02-JUL-1999  
LOCUS AF083550 Mus musculus exostotin EX111 mRNA, complete cds.  
DEFINITION AF083550  
ACCESSION AF083550.1 GI:5326768  
VERSION  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.







TITLE Okamoto, H.  
Identification of a receptor for reg (regenerating gene) protein, a  
pancreatic beta-cell regeneration factor  
JOURNAL J. Biol. Chem. 275 (15), 10723-10726 (2000)  
MEDLINE 20219101  
REFERENCE 2 (bases 1 to 3198)  
AUTHORS Okamoto, H. and Takasawa, S.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1999) Shin Takasawa, Tohoku University Graduate  
School of Medicine, Department of Biochemistry, 2-1 Seiry-machi,  
Aoba-ku, Sendai, Miyagi 980-8575, Japan  
(E-mail: shintaksw@mail.cc.tohoku.ac.jp, Tel: +81-22-717-8080,  
Fax: +81-22-717-8083)

FEATURES  
Source Location/Qualifiers  
1. 3198  
/organism="Rattus norvegicus"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/sex="male"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="adult"  
5. 2764  
/gene="Reg"  
5. 2764  
/gene="Reg"  
/codon\_start=1  
/product="Reg receptor"  
/protein\_id="BA092895.1"  
/db\_xref="GI:7339684"  
/translation="MPCYTLRANGVNGGCTCRMSNRIRLFWLSFTLPIIVFP  
LIAHYLTLDADBAKRTGPRAGNLCVKNVLDICRRESVSELDLEAKROE  
LNSIAKLNLKTEAKRSIENAKODLLOLKNVISOEHSYELAAQNPRLRL  
PERKADAGLPKPRVGRGLHNCDFSRPLSGPFEVYDSDOFAGYIDPLKOF  
QATRVANVYENAIACLYVYVGEIOEPVLDPADLEKOLHPHMRGHNHVI  
NLSKSDTONLNVSTGRAMVAOSTFYAAOYRAGPOLYSPVPHASSEPMEIPPO  
VYRKLYTFPOGKESLSRSLOEASPEEMGDDPADDDNIATLVAQDSKLD  
OVIIEPTCKNPKSLPTENALCGERDRLELTLSTFALITGDSLSLSSQCAR  
LFEALRYGAPVYVGEVQVLEPHDMQMNENALVPRVTEVHFLRSLSDSLAM  
RROGRFMETVETADSIIFNVLMIRTRIOIPAPVREVAAPFPHRSGAAGTDN  
MADGDLDPVETEPYASPKYLRFNLTIOIPAPVREVAAPFPHRSGAAGTDN  
SEAKFLSGTGFRPIGGAGSGKEFOALAGNOROPFVMTTEREYLMASLER  
LNGLPYLVNRYVWNSPKLPSEDLMPDIVPIVYTERKNSLNNRFLPNNELTEAI  
LSTDDAHLRHDELMGFRVWRKRDIVGPGRYHAMDI.PHOSWLNYSVSELSAV  
LTSAPFHKYIATLYSTMPOALRDWVDEINCEDIAMNLVSHITRKPIKYSMT  
FRCPGCPQALSHDSHERHKCINFEVKVGYMPLLYTQFRVDSVLFKRLPHDKK  
CFKFI"

gene  
CDS  
BASE COUNT 723 a 882 c 850 g 743 t  
ORIGIN

Query Match 38.0%; Score 2347.4; DB 10; Length 3198;  
Best Local Similarity 85.3%; Pred. No. 0;  
Matches 2721; Conservative 0; Mismatches 416; Indels 54; Gaps 7;

QY 590 ACATCATGAGGCGTATACATGCGGCGGAGTGGGGCGCGGGAACGAGAGTCAGACCTG 649  
Db 1 ACATCATGAGGCGTATACATGCGGCGGAGTGGGGCGCGGGAACGAGAGTCAGACCTG 60  
QY 650 CATGCTGCGCTGTCCAAACCGCATCCGCTCAGCTGAGCTCAGCTTACGCTCTTTGTCAT 709  
Db 61 TATGCTGCGCTGTCCAAACCGCATCCGCTCAGCTGAGCTCAGCTTACGCTCTTTGTCAT 120  
QY 710 CCGGCTCTTCTCCGCTCATCGCCCATATTAATCTACACACTCTGATGAGGAGCTGATGA 769  
Db 121 CCGGCTCTTCTCCGCTCATCGCCCATATTAATCTACACACTCTGATGAGGAGCTGATGA 180  
QY 770 GCGAGCAAGCGGATTTTGGTCCCGGGGGGGAACGAGCTGTGCGAGGTGAAGACAGT 829  
Db 181 GCGAGCAAGCGGATTTTGGTCCCGGGGGGGAACGAGCTGTGCGAGGTGAAGACAGT 240  
QY 830 GGTGATCTGTGCGGATCCGGGAGTGTGAGTGAAGAGCTCTGAGAGTGAAGCCAA 889  
Db 241 CCTAGATCTTGTGCGGATCCGGGAGTGTGAGTGAAGAGCTCTTGAAGTGAAGCCAA 300

QY 890 GCGCAAGAGCTGAACAGCGAGATCGCCAAAGCTGAATCTGAAGTCAAGCTGTAAAG 949  
Db 301 GCGCAGAGAGCTGAACAGCGAGATCGCCAAAGCTGAATCTGAAGTCAAGCTGTAAAG 360  
QY 950 GAGATGAGAGAGCGCAAGAGAGAGCTGCTCCAGCTCAAGATGTCATACCGACAGCA 1009  
Db 361 GAGATGAGAGAGCGCAAGAGAGAGCTGCTCCAGCTCAAGATGTCATACCGACAGCA 420  
QY 1010 GCATCTCTACAAAGAGCTCAATGAGCCAGAGACCGCAACCTGCTCCGCTCCGAGCT 1069  
Db 421 GCATCTCTACAAAGAGCTCAATGAGCCAGAGACCGCAACCTGCTCCGCTCCGAGCT 480  
QY 1070 GCTCCCAAG 1129  
Db 481 GCTCCCAAG 540  
QY 1130 ACACAACTGCTTGTGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189  
Db 541 ACACAACTGCTTGTGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
QY 1190 TGACAGTGAACAGATTTGCTTGGGAGCTACCTGGATCCCTGCTGCTGCTGCTGCT 1249  
Db 601 TGACAGTGAACAGATTTGCTTGGGAGCTACCTGGATCCCTGCTGCTGCTGCTGCT 660  
QY 1250 GCGCAG 1309  
Db 661 GCGCAG 720  
QY 1310 GATTAAGTGGGAG 1369  
Db 721 GATTAAGTGGGAG 780  
QY 1370 GTTGTATTCCTGCGCACACTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429  
Db 781 GTTGTATTCCTGCGCACACTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
QY 1430 ACGTAAGTCAATACACAGAACTCTCTATTAAGCTGAGTACTGAGCTGAGTCAATCT 1489  
Db 841 ACGTAAGTCAATACACAGAACTCTCTATTAAGCTGAGTACTGAGTCAATCTGAGT 900  
QY 1490 CCGAGTCCACTTCTACACTGTCAGAGTACAGAGCTGAGTCTGAGTCTGAGTCAAT 1549  
Db 901 CCGAGTCCACTTCTACACTGTCAGAGTACAGAGCTGAGTCTGAGTCTGAGTCAAT 960  
QY 1550 GGTCCATGCGCATGTCTGAGAGCCCACTCATGGAATCCACACAGAGAGAGAGAG 1609  
Db 961 GGTCCATGCGCATGTCTGAGAGCCCACTCATGGAATCCACACAGAGAGAGAGAGAG 1020  
QY 1610 GAATATCTCTTCACTTCCAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669  
Db 1021 GAATATCTCTTCACTTCCAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1670 GGGCCGCTCTTGAAG 1729  
Db 1081 GGGCCGCTCTTGAAG 1140  
QY 1730 CATTCGACCCCTGAAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789  
Db 1141 CATTCGACCCCTGAAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1790 CTGCAAAAACCAAGCCCAAAACCAAGCTGCGGAGTGAAGTGGGAGAGAGAGAG 1849  
Db 1201 CTGCAAAAACCAAGCCCAAAACCAAGCTGCGGAGTGAAGTGGGAGAGAGAGAGAG 1260  
QY 1850 GGAACCGCTTGAATTTGCTGAGAGTCCACACTGCGCCATCATCATTAACCCCGGGAG 1909  
Db 1261 GGAACCGCTTGAATTTGCTGAGAGTCCACACTGCGCCATCATCATTAACCCCGGGAG 1320  
QY 1910 TCGCTTGTATTTCTCTGAGGAGTGAACACAGAGCTCTTGAAGGCGAGAGAGTGGTGC 1969  
Db 1321 TCGCTTGTATTTCTCTGAGGAGTGAACACAGAGCTCTTGAAGGCGAGAGAGTGGTGC 1380



QY	1970	CGTCCGGTGTGCTGGGGGAGCAGGTCACGCTTCCCTACCGAGCAATGTCTGACGTGGAA	2029
Db	1381	TGTGCTGTGTTCCTTGGGAGCAGGTGACGCTCCGTAACGACGACATGTACAAATGGAA	1440
QY	2030	CGAGGCGGCCCCGTGGTGGTGCAAAGCCTGTGTACCGAGGTTCAATTTCCTGTCCAGAG	2089
Db	1441	TGAGGCGCCCTTATGGTGTGCCAAAGCCTGTGTACAGAGGTTCACTTCTGTGTACGAG	1500
QY	2090	CCCTCCGATAGTACCTCCCTGGCTATGAGAGCGGCAGAGCCCTTCTCTGGAGACTTA	2149
Db	1501	TCTGTCAAGACAGTGTATCTGTGTGGCTATGAGAGCGGCAGAGCCCTTCTCTGGAGACTTA	1560
QY	2150	CTTCTCCATGTCTGACAGTATTTTTAATACCGTGTGCTGCTATGATTTAGAGACTCGCATCA	2209
Db	1561	CTTCTCCACCCCTGACAGTATTTTTAATACCGTGTGCTGCTATGATTTAGAGACTCGCAATTC	1620
QY	2210	GATCCCAAGCCGCTCCCATCCGGGGAAGAGCGGAGCTGAGATCCCCACCGTTTCAAGGGA	2269
Db	1621	GATCCCAAGCTGTCTCCCATCCGGGGAAGAGGATGAGATCCCCCATCTGTTCAGGGGA	1680
Db	2270	GGCGGCTGGAACTGACCCCAACATGTGCTGCAACAGGGGACCTGAGACTGGGGCCAGTGG	2329
Db	1681	GGCAGCTGTACTATGACCCCAACATGTGCTGCAATGGGGACCTGAGACTGGGGCCGCTAGA	1740
QY	2330	GACGAGGCCGCTTACGCCCTTCCACGAGATACCTCGGCAATTTCTACTGTGACTCTGA	2389
Db	1741	GACAGAGCCGCTTATGCTCTACCTAAATACCTCCGTAATTTCACTGTGACTGTACATGA	1800
QY	2390	CTTTTACCGGACGTGGAAACGTGTCTCAGGGCCCTTCCATCTTTTCCGCCACATCCCTT	2449
Db	1801	CTTTTACCGGACGTGGAAACGTGTCTCAGGGCCCTTCCATCTTTTCCACACATCCCTT	1860
QY	2450	TGACCTGTGTGTGCCCTACAGAGGCCAAATTTCTTGGGCTCAGAGGACTGTGTTCGGCTAT	2509
Db	1861	TGACCTGTGTGTGCCCTCCTAGAGGCCAAATTTCTTGGGCTCAGAGGACTGTGTTCGGCCAT	1920
QY	2510	TGTGTGTGAGCTGGGGCTTCTGGCAAGAAATTTACGACGAGCGTTGGAGGCATGTTCC	2569
Db	1921	CGGTGTGTGGGCTGGGGGCTCTGGCAAGAGTTTCCAGGACGCGCTTGGAGGCATGTTCCA	1980
QY	2570	CCGAGAGCATTCACGGTGGTGATGTTGACTTATGAGCCGGGAGGAAGTGTATGACATC	2629
Db	1981	CGGGAGGATTTCACTATTTGATGTCTGACATTCAGAGCCGGGAGGAAGTGTATGACATC	2040
QY	2630	TTTATAGAGAGCTGATAGGCTTCCCTTACCTGACGAAGAGTGTGTGTGGATGATTTCTC	2689
Db	2041	CCTGAGAGGCTCAATGGCTCCCTTACCTGACGAAGAGTGTGTGTGGATGATTTCTCC	2100
Db	2690	CNAGCTGCCATCAGAGGACTTCTGTGGCTGACATTTGGCGTTTCCCATATGCTGTCCG	2749
Db	2101	CNAGCTGCCCTCGGAGGAGACCTTTTGTGTGGCAGACATTTGGTGTCCCATATGATGTTGTCCG	2160
QY	2750	TATCTGAGAAAGAAAGTTTGAACAACCGATTTCTTACCCTGGATGAAATTTGAGACGAGGC	2809
Db	2161	TATCTGAGAAAGAAAGTTTGAACAATCGGTTCTTGGCTTGGATGAGATGAGACGAGAGC	2220
QY	2810	CATCTGTCCATTTATGACATGCTACACTCCGCACTGACGAAGAAATCATGTTGGGTTCCG	2869
Db	2221	AATATATGTTCATTCATATACGATGTCCCACTCCGCACTGTATGAAATCATGTTCGGGTTCCG	2280
QY	2870	GGTGTGAGAGAACGCTCGGGAGCGCATGTGTGGCTTCCCTGGCGCTTACACGCAATGGGA	2929
Db	2281	GGTGTGAGAGAGCGCGGTATGCGATGTGTGGGCTTCCCTGGCGCGTACATGCGTGGGA	2340
QY	2930	CATCCCCCATATAGTTCGTGGCTCTTACAACTTCGAATCTCTGTGAGCTGTCCATGTGTCT	2989
Db	2341	CATCCCCCATACCAATGTCGTGGCTCTTACAACTTCGAATCTCTGTGAGCTGTCCATGTGTCT	2400
QY	2990	GACAGGTCTCTCTTCTTTTACAAGATTTATGTGCTACCTGTATTTCTTATGTGATGATCCCA	3049
Db	2401	GACGGGTCTCTCTTCTTTTACAAGATTTATGTGCTACCTGTATTTCTTATGTGATGATCCCA	2460
QY	3050	GGCCATCCGGGACATGTGTGATGATATACATCACTGTGAGCAATTTGCCATGATCTTCT	3109

[illegible]



D	235784	CAGAAACTTGTGTGTGAAA	-AGAGGTTGCGTTTGTGACAGACATCATGTTATGGCGA	235726
Q	554	GTGACCCGACGCTGATCTTG	--GGGGGAGGCGTGCAGAGAGCATGACAGGCTATACAT	610
D	235725	GGGATCCGACGCTGATCAGAGTGGGCAAGAGGACACAGGAACTCATGACAGCTATACAT		235666
Q	611	GCTGCGGAAATGGGGGCGGGGAAACGAGAGTCAAGACTGCTGCGCTGTGTCAACG		670
D	235665	GTTGGGGAATGGGGGAGTGGGGAAGGGTGTGCACAGCTGTATGCTGGCGTGTCAATG		235606
Q	671	CATCCGCGCTCAAGTGGGCTCAGGCTTCAGGCTCTTGTATCATCTGGTCTTCTTCCCGCAT		730
D	235605	CATCCGCGCTGACATGGCTGAGATTTACAGGCTTTCATATCATCTGCTGCTTCTTCCCGCAT		235546
Q	731	CGCCCACTATTTACCTCACCACCTCTGGATGAGCTGATGAGGCAAGCGGATTTTGG		790
D	235545	TGCTCATATTTACCTCACCACCTCTGGAGAGGCAAGAGGAGGAGGAGGATCTTGG		235486
Q	791	TCCCGGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACTGCTGGATGTGTGCCGATCG		850
D	235485	CCCTTGGGGCTGGCAGTGGCTCTGTGAGTAAAGCACGCTCGATCTGTGTGGATTGC		235426
Q	851	GGAGTCGGTGAAGTGAAGGCTCTGACAGCTGGAGGCGCAAGCGCAAGCTGAACGCGA		910
D	235425	TGAGCTGTGTGACGCAAGAGCTTGTACAGCTGGAAGCCAAAGGAGAGGCTGAACGCGA		235366
Q	911	GATCGCCAAAGCTGATCGAATCGAAGGCTGTAAAGAAAGCAATTTAGAAACCCCAAGCA		970
D	235365	GATTGCGAAGCTGAAACCTCAAGATTTGAAGGCTGTAAAGAGCATAGAGAAATGCGCAAGCA		235306
Q	971	GGACCTGCTCAGCTCAAGAAATGTATCAGCGACAGCCAGACATTTCTTCAAGAGCTCAT		1030
D	235305	GGACCTGCTGACGCTCAAGAAATGTATCATTAGCAACAGACACATCTTCAAGAGGCTGAT		235246
Q	1031	GGCCCAAGAAACGACCCCAAGCTGTGCTCCATTCGCACTGCTGCCAAGAAAGACATGC		1090
D	235245	GGCCCAAGAAACGACCCCAAGCTGTGCTCCATTCGCACTGCTGCCAAGAAAGACATGC		235186
Q	1091	GGGAC -TCCCTTCCCGCAAGGCGCACTGGGGCTCCCGGCTACACAATGCTTTGATTATT		1149
D	235185	CGGCTTCCCAACCCCGCAAGGCTCACTGGGGTTCGCCCTTCAACAATGCTTTGATTACT		235126
Q	1150	CTGCTGTGCCCTCTCAGCTCTGTGGTCCCGGCTACAGTCTATGACAGTGAACAGTTGTCT		1209
D	235125	CTGCTGTGCCCTCTCAGCTCTGTGGTCCCGGCTACAGTCTATGACAGTGAACAGTTGTCT		235066
Q	1210	TTTGGACGCTACCTGGATCCCTTTGGTCAAGGAGCTTTTCAAGGCGACAGCAAGCTTAAG		1269
D	235065	TTTGGACGCTACCTGGACCTTTTGGTCAAGGAGGCTTTTCAAGGCGTGAAGGCGCAAGC		235006
Q	1270	TTTATGTTACAGAAATTCAGACATGCGGCTCCCTTTCAGGATCTATGTTGGGAGAGATGC		1329
D	235005	TTTATGTTACAGAAATTCAGGCGCATCGGCTGCTGTATGTGTGTATGTGTGGGAAATGC		234946
Q	1330	AGGAGCCCGGTGTGCTGCGGCTCTGTAAGTGAAGAAAGTTGTATTCCCTGCCACACT		1389
D	234945	AGGAGCCCGCATGTGCTGCGGCTGCCGACCTTGAAGAAAGAGCTGTTTCTCTGCCACACT		234886
Q	1390	GGCGGACGATGACACACCATGTCTATCATTAATCTGTCAAGTAACTAGTACAGTACAGA		1449
D	234885	GGAGGACGATGAGGCGACAAACGCTCATTTATCAACGTCGTCCCGAACTCAGACACAGA		234826
Q	1450	ACCTTCTCTTAAAGTCACTACTGGCCGTGCATGGTGGGCGCAAGTCCACCTTCAACCTG		1509
D	234825	ATCTTCTCTTAAAGTCACTACTGAAGCCGGGCGGCGGATGTGGCCCACTCATATGCTG		234766
Q	1510	TTCAGTACAGACTGGCTTTGACTTGTGTCTATACCGCTGTGTCATAGCCATGTGAGC		1569
D	234765	CCCAATACAGAGCTGTGCTTTACCTGCGTGTACCCCTTGTCTCATGATATGTCTGAAC		234706
Q	1570	CCAATTCATGGAATTCACCACACAGGTGCCGCTGAAGCGGAATATCTTCACTTCC		1629

Db 234705 CCGACCTCATGGAAATCCACCGCAGGAGCGCAGTTAAGCGGAAATATCTCTCACATTTC 234646

QY 1630 AGGGCGAGAGAATTTGAGTCTCTGAGGCTTCAAGCTTTCAGAGAGGCCGCTCTTTCGAAAGAG 1689

Db 234645 AGGGCGAGAGAATATCGAGTCTCTGAGATCTTACGCTTTCAGAGAGGCCGCTCTTTCGAAAGAG 234586

QY 1680 AATATGAGGGCGGACCTCTCCGCCACTTACGATGACCGGATCATTTGGCACCTCTGAAGCGG 1749

Db 234585 AGATGGAAGGGCGACCTCTCCGCCCACTATGACGATGCTCCATTCATTTGGCCACCCCTAAAGCGC 234526

QY 1780 TGCAGAGCAGCAGCTGGATCAGGTCCTGGTGGAAATTCACATCT 1792

Db 234525 TACAGGACAGCAGCTGGATCAGGTCGCTGGTAGAATTTCCGGG 234483

LOCUS	AC011641	64626 bp	DNA	linear	HTG 29-Dec-2000
RESULT 15	AC011641/c				
DEFINITION	Homo sapiens clone RP11-13N16, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC011641				
VERSION	AC011641.2 GI:11995638				
KEYWORDS	HTG; HTGS_PHASE0.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 64626)				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens, clone RP11-13N16				
REFERENCE	2 (bases 1 to 64626)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Becker,J.R., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrelano,K., Dewar,K., Domino,M., Doneelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karkas,A., Klein,J., Lechoczky,J., Lieu,C., Locke,R., Macdonald,P., Marquis,N., Mcwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Sevely,P., Steinge-Thoman,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-Oct-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Dec 29, 2000 this sequence version replaced g1:6016773. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>				
	----- Genome Center				
	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: WIBR				
	Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a>				
	Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a>				
	----- Project Information				
	Center project name: L3369				
	Center clone name: 13_N_16				
	-----				
	* NOTE: This record contains 78 individual				
	* sequencing reads that have not been assembled into				
	* contigs. Runs of N are used to separate the reads				
	* and the order in which they appear is completely				
	* arbitrary. Low-pass sequence sampling is useful for				
	* identifying clones that may be gene-rich and allows				
	* overlap relationships among clones to be deduced.				
	* However, it should not be assumed that this clone				
	* will be sequenced to completion. In the event that				
	* the record is updated, the accession number will				
	* be replaced.				
	1 734: contig of 734 bp in length				

```
* 735 834: gap of 100 bp
* 835 1573: contig of 739 bp in length
* 1574 1673: gap of 100 bp
* 1674 2401: contig of 728 bp in length
* 2402 2501: gap of 100 bp
* 2502 3211: contig of 710 bp in length
* 3212 3311: gap of 100 bp
* 3312 4048: contig of 737 bp in length
* 4049 4148: gap of 100 bp
* 4149 4850: contig of 702 bp in length
* 4851 4950: gap of 100 bp
* 4951 5675: contig of 725 bp in length
* 5676 5775: gap of 100 bp
* 5776 6464: contig of 689 bp in length
* 6465 6564: gap of 100 bp
* 6565 7266: contig of 702 bp in length
* 7267 7366: gap of 100 bp
* 7367 8098: contig of 732 bp in length
* 8099 8198: gap of 100 bp
* 8199 8908: contig of 710 bp in length
* 8909 9008: gap of 100 bp
* 9009 9733: contig of 725 bp in length
* 9734 9833: gap of 100 bp
* 9834 10557: contig of 724 bp in length
* 10558 10657: gap of 100 bp
* 10658 11380: contig of 723 bp in length
* 11381 11480: gap of 100 bp
* 11481 12202: contig of 722 bp in length
* 12203 12302: gap of 100 bp
* 12303 13058: contig of 756 bp in length
* 13059 13158: gap of 100 bp
* 13159 13902: contig of 744 bp in length
* 13903 14002: gap of 100 bp
* 14003 14723: contig of 721 bp in length
* 14724 14823: gap of 100 bp
* 14824 15543: contig of 726 bp in length
* 15550 15649: gap of 100 bp
* 15650 16389: contig of 740 bp in length
* 16390 16489: gap of 100 bp
* 16490 17192: contig of 703 bp in length
* 17193 17292: gap of 100 bp
* 17293 18006: contig of 714 bp in length
* 18007 18106: gap of 100 bp
* 18107 18812: contig of 706 bp in length
* 18813 18912: gap of 100 bp
* 18913 19655: contig of 743 bp in length
* 19656 19755: gap of 100 bp
* 19756 20490: contig of 735 bp in length
* 20491 20590: gap of 100 bp
* 20591 21335: contig of 745 bp in length
* 21336 21435: gap of 100 bp
* 21436 22140: contig of 705 bp in length
* 22141 22240: gap of 100 bp
* 22241 22963: contig of 723 bp in length
* 22964 23063: gap of 100 bp
* 23064 23780: contig of 717 bp in length
* 23781 23880: gap of 100 bp
* 23881 24616: contig of 736 bp in length
* 24617 24716: gap of 100 bp
* 24717 25484: contig of 768 bp in length
* 25485 25584: gap of 100 bp
* 25585 26329: contig of 745 bp in length
* 26330 26429: gap of 100 bp
* 26430 27188: contig of 759 bp in length
* 27189 27288: gap of 100 bp
* 27289 28037: contig of 749 bp in length
* 28038 28137: gap of 100 bp
* 28138 28886: contig of 749 bp in length
* 28887 28986: gap of 100 bp
* 28987 29720: contig of 734 bp in length
* 29721 29820: gap of 100 bp
* 29821 30566: contig of 746 bp in length
* 30567 30666: gap of 100 bp

* 30667 31409: contig of 743 bp in length
* 31410 31509: gap of 100 bp
* 31510 32229: contig of 720 bp in length
* 32230 32329: gap of 100 bp
* 32330 33057: contig of 728 bp in length
* 33058 33157: gap of 100 bp
* 33158 33656: contig of 699 bp in length
* 33657 33956: gap of 100 bp
* 33957 34683: contig of 727 bp in length
* 34684 34783: gap of 100 bp
* 34784 35513: contig of 730 bp in length
* 35514 35613: gap of 100 bp
* 35614 36344: contig of 731 bp in length
* 36345 36444: gap of 100 bp
* 36445 37202: contig of 758 bp in length
* 37203 37302: gap of 100 bp
* 37303 38042: contig of 740 bp in length
* 38043 38142: gap of 100 bp
* 38143 38869: contig of 727 bp in length
* 38870 38969: gap of 100 bp
* 38970 39724: contig of 755 bp in length
* 39725 39824: gap of 100 bp
* 39825 40595: contig of 771 bp in length
* 40596 40695: gap of 100 bp
* 40696 41435: contig of 740 bp in length
* 41436 41535: gap of 100 bp
* 41536 42234: contig of 699 bp in length
* 42235 42334: gap of 100 bp
* 42335 43069: contig of 735 bp in length
* 43070 43169: gap of 100 bp
* 43170 43878: contig of 709 bp in length
* 43879 43978: gap of 100 bp
* 43979 44708: contig of 730 bp in length
* 44709 44808: gap of 100 bp
* 44809 45526: contig of 718 bp in length
* 45527 45626: gap of 100 bp
* 45627 46349: contig of 723 bp in length
* 46350 46449: gap of 100 bp
* 46450 47182: contig of 733 bp in length
* 47183 47282: gap of 100 bp
* 47283 48007: contig of 725 bp in length
* 48008 48107: gap of 100 bp
* 48108 48827: contig of 720 bp in length
* 48828 48927: gap of 100 bp
* 48928 49651: contig of 724 bp in length
* 49652 49751: gap of 100 bp
* 49752 50470: contig of 719 bp in length
* 50471 50570: gap of 100 bp
* 50571 51314: contig of 744 bp in length
* 51315 51414: gap of 100 bp
* 51415 52148: contig of 734 bp in length
* 52149 52248: gap of 100 bp
* 52249 52982: contig of 734 bp in length
* 52983 53082: gap of 100 bp
* 53083 53779: contig of 697 bp in length
* 53780 53879: gap of 100 bp
* 53880 54593: contig of 714 bp in length
* 54594 54693: gap of 100 bp
* 54694 55420: contig of 727 bp in length
* 55421 55520: gap of 100 bp
* 55521 56249: contig of 729 bp in length
* 56250 56349: gap of 100 bp
* 56350 57095: contig of 746 bp in length
* 57096 57195: gap of 100 bp
* 57196 57942: contig of 747 bp in length
* 57943 58042: gap of 100 bp
* 58043 58771: contig of 729 bp in length
* 58772 58871: gap of 100 bp
* 58872 59586: contig of 715 bp in length
* 59587 59686: gap of 100 bp
* 59687 60408: contig of 722 bp in length

10.1%; Score 621; DB 2; Length 64626;
```

Best Local Similarity 84.7%; Pred. No. 6,9e-136;  
Matches 651; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

```

OY 1215 AGCTACCTGGATCCCTGGTCAAGCAGAGCTTTTCAGGCGACAGACAGCTAACGTTTAT 1274
DB 60510 ACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60451
OY 1275 GTTACAGAAATGACAGATCGGCTGCTTACGTACTAGTGGAGAGATGACAGAG 1334
DB 60450 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60391
OY 1335 CCCGTGGTCTGGCGGCTGCTGAGCTGAGAGAGAGTGTATTCCCTGCCACTGGCGG 1394
DB 60390 CCGGTGGTCTGGCGGCTGCTGAGAGAGAGTGTATTCCCTGCCACTGGCGG 60331
OY 1395 ACGGATGACACAAACATGTCATCATCTGTACAGTAAGTACAGACAGACCTT 1454
DB 60330 ACGGATGACACAAACATGTCATCATCTGTACAGTAAGTACAGACAGACCTT 60271
OY 1455 CTCTATACGTCAGTACTGGCCGCTGATGAGGCGCCAGTCCACCTTCTACACTGTCCAG 1514
DB 60270 CTCTATACGTCAGTACTGGCCGCTGATGAGGCGCCAGTCCACCTTCTACACTGTCCAG 60211
OY 1515 TACAGACCTGGCTTTGACTGGTGTATACACCGCTGTCCATGCGCATGTGAGCCCAAC 1574
DB 60210 TACAGACCTGGCTTTGACTGGTGTATACACCGCTGTCCATGCGCATGTGAGCCCAAC 60151
OY 1575 TTCTATGAAATCCACACAGAGTGGCGGTGAAGGAAATATCTCTACCTTCCAGGGC 1634
DB 60150 TTCTATGAAATCCACACAGAGTGGCGGTGAAGGAAATATCTCTACCTTCCAGGGC 60091
OY 1635 GAGAAGATTGAGTCTGAGGTCTAGGCTTACAGAGGCGCGCTCTTGAAGAGAAATG 1694
DB 60090 GAGAAGATTGAGTCTGAGGTCTAGGCTTACAGAGGCGCGCTCTTGAAGAGAAATG 60031
OY 1695 GAGGCGGACCTCCCGCGGACTACGATGACCGGATCATTTGCCACCTGGAAGCGGTGAG 1754
DB 60030 GAGGCGGACCTCCCGCGGACTACGATGACCGGATCATTTGCCACCTGGAAGCGGTGAG 59971
OY 1755 GACAGCAAGCTGATCAGGTCTGTTGGAATTCACCTGCAAAAACAGCCCAACCCAGC 1814
DB 59970 GACAGCAAGCTGATCAGGTCTGTTGGAATTCACCTGCAAAAACAGCCCAACCCAGC 59911
OY 1815 CTGCGGACTGAGTGGGCACTGTGTGAGAGCGGAGGACCGCTTGAATTGCTGAAGCTC 1874
DB 59910 CTGCGGACTGAGTGGGCACTGTGTGAGAGCGGAGGACCGCTTGAATTGCTGAAGCTC 59851
OY 1875 TTCACCTTTGGCCCTCATCATTTACCCCGGGGACCTGCTTGGTTATTCTCTGGGTGT 1934
DB 59850 TTCACCTTTGGCCCTCATCATTTACCCCGGGGACCTGCTTGGTTATTCTCTGGGTGT 59791
OY 1935 GCAACAC---GGCTCTTCGAAAGC--CTGGAAGTGGGTGCGCTCCCGGTG 1979
DB 59790 GTGCAACACGGGCTTTTCGAAAGCAGATGGAAGTGGGTGCGGCGGTG 59742

```

Search completed: September 8, 2002, 23:03:16  
Job time: 18222 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: September 4, 2002, 14:58:36 ; Search time 15.05 Seconds  
(without alignments)  
2364.337 Million cell updates/sec

Title: US-09-809-920-4  
Perfect score: 4873  
Sequence: 1 MTGYTWLRNGAGNGGTCM.....DSVLEKTRLPDHDKCKEKF 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4873	100.0	919	EXL3_HUMAN	043909 homo sapien
2	1251.5	25.7	814	EXT2_MOUSE	001705 caenorhabd
3	1004	20.6	718	EXT2_MOUSE	P70428 mus musculi
4	1000	20.5	718	EXT2_MOUSE	Q33063 homo sapien
5	871.5	17.9	746	EXT1_MOUSE	P97464 mus musculi
6	871	17.9	746	EXT1_MOUSE	Q16394 homo sapien
7	660.5	13.6	676	EXT1_MOUSE	Q92936 homo sapien
8	367	7.5	378	EXT1_MOUSE	Q01704 caenorhabd
9	251.5	5.2	378	EXT1_MOUSE	Q01704 caenorhabd
10	123.5	2.5	622	VP27_YEAST	P40343 saccharomyc
11	113	2.3	676	VATI_ARCFU	Q29106 archaeoglob
12	113	2.3	2663	CENE_HUMAN	Q02224 homo sapien
13	112.5	2.3	4568	DHIC_CAEEL	Q19020 caenorhabd
14	111.5	2.3	3075	LMAL_BRAJA	P25391 homo sapien
15	110.5	2.3	429	CPXR_BRAJA	Q59264 bradyrhizob
16	110.5	2.3	1940	MYH3_CHICK	P02565 gallus gall
17	109	2.2	978	RA50_AQUAE	O67124 aquiflex aeo
18	109	2.2	4967	RYR2_HUMAN	Q92736 homo sapien
19	108.5	2.2	1311	FMR2_HUMAN	P51816 homo sapien
20	108.5	2.2	1595	SOS_DROME	P26675 drosophila
21	108	2.2	873	PC1_HUMAN	P22413 homo sapien
22	107.5	2.2	1087	ON53_HUMAN	Q95683 homo sapien
23	107	2.2	1748	POLR_ELV	P35928 erysimum la
24	107	2.2	4684	PLE1_HUMAN	Q15149 homo sapien
25	106.5	2.2	700	HS9C_DICDI	P54651 dictyosteli
26	106.5	2.2	2324	COAC_CHICK	P11029 gallus gall
27	106	2.2	732	ATZN_ECOLI	P37617 escherichia
28	105.5	2.2	1976	MYHA_BOVIN	Q27991 bos taurus
29	105.5	2.2	2871	DESP_HUMAN	P15924 homo sapien
30	105	2.2	474	YLEA_HAEIN	Q57163 haemophilus
31	104.5	2.1	1505	CUT2_HUMAN	O14529 homo sapien
32	104.5	2.1	4128	PRKD_HUMAN	P78527 homo sapien
33	104	2.1	359	PEXC_CAEEL	Q19189 caenorhabd

ALIGNMENTS

RESULT	1	STANDARD	PRT	919 AA.	
EXL3_HUMAN					
ID	EXL3_HUMAN				
AC	043909; 000225;				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Exostosin-like 3 (putative tumour suppressor protein EXTL3) (multiple exostosin-like protein 3) (Hereditary multiple exostososes gene Isolog)				
DE	(EXT-related protein 1).				
GN	EXTL3 OR EXTL1 OR EXTL2 OR KIAA0519.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI TaxID:9606;				
RA	11				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-98140122; PubMed-9479495;				
RA	Ivan Hul W., Wuyts W., Hendrickx J., Speleman F., Wauters J.,				
RA	de Boule K., van Roy N., Bossuyt P., Williams P.J.;				
RT	"Identification of a third EXT-like gene (EXTL3) belonging to the EXT				
RT	gene family.";				
RL	Genomics 47:230-237(1998).				
RN	121				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-Testis, and Fetal;				
RX	MEDLINE-98139867; PubMed-9473480;				
RA	Sato T., Seki N., Yamauchi M., Tsuji S., Hayashi A., Kozuma S.,				
RA	Hori T. A.;				
RT	"Structure, chromosomal location, and expression profile of EXTL1 and				
RT	EXTL2, new members of the multiple exostososes gene family.";				
RL	Biochem. Biophys. Res. Commun. 243:61-66(1998).				
RN	131				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-Brain;				
RX	MEDLINE-98290545; PubMed-9628581;				
RA	Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,				
RA	Nomura N., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. IX.				
RT	The complete sequences of 100 new cDNA clones from brain which can				
RL	code for large proteins"-in vitro.";				
RL	DNA Res. 5:31-39(1998)				
RN	141				
RP	SEQUENCE FROM N.A.				
RA	Deng H.-X., Xu L., Xia J.-H., Fan C., Pan Q., Liu C.-Y., Ruan Q.-G.;				
RT	"Molecular cloning of a candidate gene for multiple exostososes.";				
RL	submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
RN	151				
RP	SEQUENCE FROM N.A.				
RA	Sato T.;				
RT	"The human EXTL1/EXTL2/EXTL3 gene at 8p11-p12, the third breast				
RT	cancer susceptibility gene locus, is not mutated in breast and various				
RL	cancers.";				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RN	161				
RP	SEQUENCE OF 470-919 FROM N.A.				

34	104	2.1	797	1	SACB_STREMU	P11701 streptococc
35	104	2.1	1875	1	MLP1_YEAST	Q02455 saccharomyc
36	103.5	2.1	561	1	RECB_SCHPO	P36266 schizosacch
37	103.5	2.1	1046	1	POL_FENY1	P31792 felline endo
38	103.5	2.1	1118	1	CARB_YEAST	P03965 saccharomyc
39	103	2.1	869	1	YD95_YEAST	P38966 saccharomyc
40	103	2.1	2346	1	COAL_HUMAN	Q13085 homo sapien
41	102.5	2.1	829	1	CRCM_HUMAN	P23508 homo sapien
42	102.5	2.1	839	1	VPR1_MOUSE	Q92194 mus musculi
43	102.5	2.1	926	1	UVRA_AQUAE	O66911 aquiflex aeo
44	102.5	2.1	1002	1	ODOL_HUMAN	O02218 homo sapien
45	102.5	2.1	1006	1	K125_TORAC	O23826 nicotiana t

RA Adams M.D.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE TUMOR SUPPRESSOR (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic  
 CC reticulum (By similarity).  
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED.  
 CC -1- DISEASE: CANDIDATE GENE FOR THE BREAST CANCER LOCUS ON CHROMOSOME  
 CC 8p12-p22.  
 CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-6 OR MET-20 IS THE  
 CC INITIATOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF001690; AAC39598.1; -  
 DR EMBL: AB007042; BAA24080.1; -  
 DR EMBL: AB011091; BAA25445.1; -  
 DR EMBL: AF029231; AAD01877.1; -  
 DR EMBL: AF083551; AAD42041.1; -  
 DR EMBL: U96629; AAB67602.1; ALT\_SEQ.  
 DR MIM: 605744; -  
 DR InterPro: IPR004263; Exostosin.  
 DR Pfam: PF03016; Exostosin; 1.  
 KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.  
 FT TRANSMEM 31  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 692 695  
 FT CARBOHYD 277 277 POLY-VAL.  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 919 AA; 104748 MW; 200AD4DAB4A39FD CRC64;  
 Query Match 100.0%; Score 4873; DB 1; Length 919;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTGTYMLRNGAGNGGOTCMRLMSNRIRLWLSFTLVIIIVFFPLIAHYITLDEADEA 60  
 DB 1 MTGTYMLRNGAGNGGOTCMRLMSNRIRLWLSFTLVIIIVFFPLIAHYITLDEADEA 60  
 Y 61 GKRIFFRGVNGELCEVKNHVDLCIRISVSEBELLQLEAKROELNSETAKINLKEACKS 120  
 DB 61 GKRIFFRGVNGELCEVKNHVDLCIRISVSEBELLQLEAKROELNSETAKINLKEACKS 120  
 Y 121 IENAKODLLQKNYISOTESHYKELMANONPKLSIRLLPEKODAGLPKPKATRGRLH 180  
 DB 121 IENAKODLLQKNYISOTESHYKELMANONPKLSIRLLPEKODAGLPKPKATRGRLH 180  
 Y 121 IENAKODLLQKNYISOTESHYKELMANONPKLSIRLLPEKODAGLPKPKATRGRLH 180  
 QY 181 NCFDYSNCPILTSGFPPVYVSDQVFSGSYLDPYLKQAFQATARANYVTENADIACYVI 240  
 DB 181 NCFDYSNCPILTSGFPPVYVSDQVFSGSYLDPYLKQAFQATARANYVTENADIACYVI 240  
 Y 241 LVGEMQEPVYLRAPELEKQYSLPHMFTDGHNNYIINLSRKSPTQNLNLYVSTGRANVAQ 300  
 DB 241 LVGEMQEPVYLRAPELEKQYSLPHMFTDGHNNYIINLSRKSPTQNLNLYVSTGRANVAQ 300  
 Y 241 LVGEMQEPVYLRAPELEKQYSLPHMFTDGHNNYIINLSRKSPTQNLNLYVSTGRANVAQ 300  
 DB 301 STFTYVYRGGFDLYVSPLVHAMSEPMFMEIPQVPRKRYLFTFOGEKTESLSSQEA 360  
 QY 361 RSFEENEKGPDPADYDRITATLKAODSKLDVLEFTCKNPKSLPTFEMALCGRED 420  
 DB 361 RSFEENEKGPDPADYDRITATLKAODSKLDVLEFTCKNPKSLPTFEMALCGRED 420  
 QY 421 RLELKLSTFALITTPGDPRLVYSSGCATRLFEALEVGAVPVVLGDEVOQLPYQDMLQNE 480

DB 421 RLELKLSTFALITTPGDPRLVYSSGCATRLFEALEVGAVPVVLGDEVOQLPYQDMLQNE 480  
 QY 481 AALVPPKRVYEVHFLRSISDSPLLAMRRGCRFLMEYESTASINTVAMIRTIQI 540  
 DB 481 AALVPPKRVYEVHFLRSISDSPLLAMRRGCRFLMEYESTASINTVAMIRTIQI 540  
 QY 541 PAAPIREAAAEIHRSGKAAGTDPNNADNGDLGLGVETEPYASPRYLNFYVYTD 600  
 DB 541 PAAPIREAAAEIHRSGKAAGTDPNNADNGDLGLGVETEPYASPRYLNFYVYTD 600  
 QY 601 YRSNCAAPGPHLPHPPDPVLPSEAKFLGSGTGFPRIGGAGSGKERQALGAVPR 660  
 DB 601 YRSNCAAPGPHLPHPPDPVLPSEAKFLGSGTGFPRIGGAGSGKERQALGAVPR 660  
 QY 661 EOPVVMVLYTEREVLNLSERLGLYLKVVVWVNSPKLPSDILLMPDIGVIMVVT 720  
 DB 661 EOPVVMVLYTEREVLNLSERLGLYLKVVVWVNSPKLPSDILLMPDIGVIMVVT 720  
 QY 721 EKNLSANRFLPWNIEETALISIDDAHLRDEIMFGFRVREARDRIVGPGRYHAMDI 780  
 DB 721 EKNLSANRFLPWNIEETALISIDDAHLRDEIMFGFRVREARDRIVGPGRYHAMDI 780  
 QY 781 PHOSMLYNSVSCSLSVLGAFFHKYAYLYVMPQAIRDMVDEYINCEDIAMNELY 840  
 DB 781 PHOSMLYNSVSCSLSVLGAFFHKYAYLYVMPQAIRDMVDEYINCEDIAMNELY 840  
 QY 841 SHIRKRPPIVTSWTFRCQCPQALSHDSSHFERHKCINFPKYVGYMPLLYTORVD 900  
 DB 841 SHIRKRPPIVTSWTFRCQCPQALSHDSSHFERHKCINFPKYVGYMPLLYTORVD 900  
 QY 901 SVLEKTRLPDKTKCFKFI 919  
 DB 901 SVLEKTRLPDKTKCFKFI 919  
 RESULT 2  
 ID EXT2\_CAEEL STANDARD; PRT; 814 AA.  
 AC 001705;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE Multiple exostosins homolog 2.  
 GN RIB-2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9726342; PubMed=9110175;  
 RA Clines G.A., Ashley J.A., Shah S., Lovett M.;  
 RT "The structure of the human multiple exostosins 2 gene and  
 RT characterization of homologs in mouse and Caenorhabditis elegans.";  
 RL Genome Res. 7:359-367(1997).  
 CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U94835; AAC47510.1; -  
 DR InterPro: IPR004263; Exostosin.  
 DR Pfam: PF03016; Exostosin; 1.  
 SQ SEQUENCE 814 AA; 94182 MW; 0A9A2AC7CE308E14 CRC64;  
 Query Match 25.7%; Score 1251.5; DB 1; Length 814;



```

QY 356 SLQARSFEEMEGDPPADYDRIATATKAYODSKLDVLYEFTCKNPKPSLPTMALC 415
Db 258 QMAHPEYRELE-----ALQAKHSESVLADKCTNSEGLSVR-KRC 300
QY 416 GERE--DRLELLKLTSTFALLITPGDRLVSSGATRLFELEVGAVVYGEQQLRYQ 473
Db 301 HOHGVFDLPVQLQENTFCTVLRRA--RL-----GQAV--LSDLVQACVYVADSTILPFS 353
QY 474 DMLQWNEALVYKPRVTEVHFLRLSLSDSLAMRQGRFLMETYFSTADSIENFTVLM 533
Db 354 EVLDMKASVYVPEEKMDVYSILQNIPOROIEMQROARFWEAYFQSTAKALATLQI 413
QY 534 IRTIQIAPAIIRERAAEIPHRSKAGCTPMADNDLGLGVETEPYPASPRYLNKF 593
Db 414 INDRY-----PYAAIS----- 425
QY 594 TLFTVDFRSWNCAPGPHLPHP--PDVLPSEAKFLGSGTGRFGGAGSGSKERQA 652
b 426 -----YEENNDPPAYKMAVSNDPLFLPLIPPOS----- 453
QY 653 ALGNVPREQFTVYMLTYEREVEVLNLSERLNGLPYLKRVVYVNSP-KLPSDILLMPDI 711
Db 454 -----OGFTALVLTFRVESLFRVLTESKVPISLKLTVNNQKNPESIMPKI 505
QY 712 GVPIMVYRTKNSLNRLPNNEIETAILSDDD-AHLRDEIMFGFRVREARDRIVG 770
Db 506 RVPKRVYTAENKLSNRFPEDEIETEAVLAIDDDIMLTSDLEQFGYEVAREFPDRLVG 565
QY 771 FPGRYHAMDPFHOSMLYNSYSCELSMYLTGAFFHKYAYLYSYVMOALRDVDEYN 830
Db 566 YPGSLHMDHMKKMKKSEKTEWNEVSWLTCALYHKYFNLYTKKMGDKIKNVADAHN 625
QY 831 CEDIANNFVSHITRKPKPKVTSWTFRCGCP--QALSHDHSFHERHKCINFEVYVG 888
Db 626 CEDIANNFVANYGKAVIKVTPRKKECECTAIDGLSDQTHMVERSECINKFAVFG 685
QY 889 YMPLLYQFRYDYLKTRLPDHTKCF 916
Db 686 TMLKVEHRRADPYLRKDDP-EKLKSF 712

```

```

RT exostoses. ";
RL Hum. Mol. Genet. 5:1547-1557(1996).
RL [3]
RP SEQUENCE FROM N.A.
RX Clines G.A., Ashley J.A., Shan S., Lovett M.;
RT "The structure of the human multiple exostoses 2 gene and
RL characterization of homologs in mouse and Caenorhabditis elegans. ";
RN Genome Res. 7:359-367(1997).
RP VARIANT EXT2 ASN-227.
RX MEDLINE-97465597; PubMed-9326317;
RA Philippe C., Porter D.E., Emerton M.E., Wells D.E., Simpson A.H.R.W.,
RT Monaco A.P.;
RT "Mutation screening of the EXT1 and EXT2 genes in patients with
RL hereditary multiple exostoses. ";
CC Am. J. Hum. Genet. 61:520-528(1997).
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum (by similarity).
CC -1- DISEASE: DEFECTS IN EXT2 ARE ASSOCIATED WITH HEREDITARY MULTIPLE
CC EXOSTOSES (EXT). EXT IS A GENETICALLY HETEROGENEOUS BONE DISORDER
CC CAUSED BY GENES SEGREGATING ON HUMAN CHROMOSOMES 8, 11, AND 19 AND
CC DESIGNATED EXT1, EXT2 AND EXT3 RESPECTIVELY. EXT IS A DOMINANTLY
CC INHERITED SKELETAL DISORDER PRIMARILY AFFECTING ENCHONDRAL BONE
CC DURING GROWTH. THE DISEASE IS CHARACTERIZED BY FORMATION OF
CC NUMEROUS CARTILAGE-CAPPED, BENIGN BONE TUMORS THAT ARE OFTEN
CC ACCOMPANIED BY SKELETAL DEFORMITIES AND SHORT STATURE. IN A SMALL
CC PERCENTAGE OF CASES EXOSTOSES HAVE EXHIBITED MALIGNANT
CC TRANSFORMATION RESULTING IN AN OSTEOSARCOMA OR CHONDROSARCOMA.
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U62740; AAB07008.1; -
DR EMBL: U64511; AAC50764.1; -
DR EMBL: U63368; AAC51219.1; -
DR EMBL: U67356; AAC51219.1; JOINED.
DR EMBL: U67357; AAC51219.1; JOINED.
DR EMBL: U67358; AAC51219.1; JOINED.
DR EMBL: U67360; AAC51219.1; JOINED.
DR EMBL: U67361; AAC51219.1; JOINED.
DR EMBL: U67362; AAC51219.1; JOINED.
DR EMBL: U67363; AAC51219.1; JOINED.
DR EMBL: U67364; AAC51219.1; JOINED.
DR EMBL: U67365; AAC51219.1; JOINED.
DR EMBL: U67366; AAC51219.1; JOINED.
DR EMBL: U67367; AAC51219.1; JOINED.
DR MIM: 133701.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
DR Anti-oncogene; Multigene family; Transmembrane; Signal-anchor;
KW Disease mutation.
FT TRANSMEM 26 46
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT CARBOHYD 288 288
FT CARBOHYD 637 637
FT VARIANT 227 227
FT D -> N (IN EXT2).
FT FTID-VAR.002378.
SQ SEQUENCE 718 AA: 82254 MW: 9048CDJA5B63C5CB CRC64;

```

Query Match 20.5%; Score 1000; DB 1; Length 718;  
 Best Local Similarity 32.5%; Pred. No. 3, 7e-64;  
 Matches 260; Conservative 125; Mismatches 258; Indels 156; Gaps 27;  
 140 HSYKEIMQNOPKLSL---PIRLPEKDAGLP-PPKATRGRLHNCFDYSRCPLTSGP- 194

```

Db 48 HSTESSNDMVEKRSTIDVYVRLP-----ADSIPIPERGDLSCNMHTCFDYRC---GFN 99
QY 195 -----PYVYVDSQFV--FGSYLDPLVYKQAFQATARA---NVVTENADIACLVILVGE 244
Db 100 PKKIKYIYALKKYVDDEFSVSNITISREYNELMALISDSOVYTTDINKACLFVPSIDV 159
QY 245 MOEPVLRPALEKOLYSLPHMRTDGHNVITN-LSRKSPTQNLVNVSGRAMVAOSTF 303
Db 160 LNQN-TLRKETQAAMQALSRM-DRGTNHLDFNMLPGPPDYNTALDVPDRALLAGGCF 217
QY 304 YTVQYRPGFGLVYSLVHNASERNFMETPPQVYKRYLFTFGGEKESLSLSQEARSF 363
Db 218 STYTVRQGVDSI-PYSPSLAE--VDLPKGGPQYFL-----LSSQGLPEY 265
QY 364 EEEEGGDPADYDRIATLTKAVODSKLDVLEFTCKNPKPSLPTFMALCGERE--DR 421
Db 266 REDELE-----ALQVKGESVLYDKCTNLSEGLSVR-KRCHKHOFDY 308
QY 422 LELIKLSTALITTPGDPRLVYSSGATRLFEALVGAVPVLYGEQVLPYDMLOWNEA 481
Db 309 PVLQGEATFCVLR--GARL---GGAV-LSQVLAQCVPVYADSYILPFESEVLDWKRA 361
QY 482 ALVYPRPVTEVHFLRLSLSDLLMRGGRFLMETYSTADSIPTVLAIRTRIQIP 541
Db 362 SVVYPERKMSDVYSILQSIQROIEMQORARFWEAYFQSIAIALATLQIINDRIY 421
QY 542 AARIREEAAEIPHRSGKAGTDPNMADNGDLGVPYETEPYASPRYLNNFTLVDFY 601
Db 422 AALSYEE-----WNPRAV----- 435
QY 602 RSNMNCAGPRLHPHPPFDVLPSEAKFLGSGFRPDIGGAGSGKEFOALGVNPRE 661
Db 436 -KMGVSNPFLF-----PLIPPOG-----Q 454
QY 662 QFVVMILTYREEVYVMSLERLNGLPRLNKVYVYVWNSP-KLPSEDLMPDQVIMVYRT 720
Db 455 GFATVLTLYRVESLFEVITEVSKVPSLSKLLVVMNQNNPPEDSLMPKIRVPLKVVRT 514
QY 721 EKSLNRLFLPMNEIETEAISIDDD-AHLRDEIMGFVFWREARIRYIGFPGRYHAM 779
Db 515 AEKLSRFRPYDEIETEAVALDIDDLIMLTSDELQCYEVMRFRPRLYGVRRLHMD 574
QY 780 IPIQSWLYNSVSCSLSVLTGAFFHKKYAYLYSYVPOAIRDMVDEYINCEDIANFL 839
Db 575 HEMNKMVEESEMTEVSMVLTGAFFHKKYAYLYTYTKMPGDIKMWDAHNMCEDIANFL 634
QY 840 VSHITRPRPKYVSRMTFRGCGP--QALSHDSHFHRKKTINFYKVVGYVMPLYTOP 897
Db 635 VAVVTGAVALTKVPRKFKCPECTALDGLSLDOTHMYERSECIINKFASVFGTMLKVEH 694
QY 898 RVDSVLEKTRLPKHDTKCF 916
Db 695 RADPVLYKKDFFP-EKLKSF 712

```

```

Rx MEDLINE-97330052; PubMed-9186511;
RA Lohmann D.R., Bulting K., Lueddecke H.J., Horsthemke B.;
RT "The murine Ext1 gene shows a high level of sequence similarity with
RT its human homologue and is part of a conserved linkage group on
RL chromosome 15";
RN Cytogenet. Cell Genet. 76:164-166(1997).
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6;
RX MEDLINE-97396026; PubMed-9254013;
RA Lin X., Wells D.;
RT "Isolation of the mouse cdna homologous to the human EXT1 gene
RT responsible for hereditary multiple exostoses";
RL DNA Seq. 7:199-202(1997).
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U78539; AABA1728.1; -.
CC EMBL: X96639; CAA65443.1; -.
CC MGD: MGI:894663; Ext1.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT TRANSMEM 8 28
FT FT
FT CARBOHYD 89 89
FT CARBOHYD 330 330
FT CONFLICT 16 16
FT CONFLICT 60 60
FT CONFLICT 220 220
FT CONFLICT 486 486
FT CONFLICT 548 548
SQ SEQUENCE 746 AA; 86307 MW; 5AC7F24BCEDEF9 CRC64;

Query Match 17.9%; Score 871.5; DB 1; Length 746;
Best Local Similarity 28.6%; Pred. No. 7e-55;
Matches 240; Conservative 142; Mismatches 245; Indels 211; Gaps 33;

QY 139 EHSYKELMAONQK---LSLPIRLLP-----EKDDAGLP-PRKATRG-----C 177
Db 39 EHSGRNGLHQPSRDHFWRRPRDALRPFPPWDOLENNEDSVHSIPROKRDANSSYYKSKC 98
QY 178 RLHNCPRYSRCPRLSGFRVVY--DSQFVFGSYLDPLVYQAFQATARAAYVYVENDIA 235
Db 99 RMESCFDEFTLCK-KNGKVVYVYPOQGEKTAEST-----QNLIAIEGSRFTSDSQ 151
QY 236 CLVYILVGEQEPVYLRPA--ELEKOLYSLPHMRTDGHNVITN---SRKSPTQNLV 289
Db 152 CLVFLSIDTLDRD-QLSQVYVHNLRSKVQSLHLM-NNGRNHLFLNLYSGWIPDVTEDVG 209
QY 290 NVSTGRAMVAOSTFYVYVYRPGFGLVYSLV---HAMSEPNFMETPPQVYKRYLFT 344
Db 210 DI--QAMLAKKASISTENFRPNFVSI-PLFSKDHPRGTGGERGLK-NTIPRLRKMYLV 265
QY 345 FQGEKI-----ESLRSSLSQARSFEEMEGDPRADYDRIATLTK-----AVODSKLD 393
Db 266 FKGRKRYLTGSDTRNLVYVHNGEDVL-----LITTCCKHGKDMQKRNHSDSCDR 314
QY 394 VIVEFTCKNQPKSLPTFMALCGEREDRLKLTSTFALITTPGDPRLVYSSGATRLPE 453
Db 315 DMEY-----EKDYREMIHNAITGCLV--PRGRRL-----GSFRFLE 349

```





[illegible]

```

QY 393 QVLEFCKNPKPSLPTMALCGEREDRLLELKLSTFALITTPGDPRLVSSG---CA 448
Db 252 -----CEODPGPS-----QTORQETLPNATCCLT-----SGHRPEAA 283
QY 449 TRLEPAAVGAAPVVLGQVOVLPRYODMLQWNEALVPRVTEVHFLRLSLSDLLAM 508
Db 284 SRLFOALQAGCIPVLISPRMELPFSEVIDWTKAIVADERLPLQVLAALQEMSPARVAL 343
QY 509 RROGRELMEYESTADSLFNIVLMIKTRIQIPAAPIEEAAAEIPHSKRAKATDPNMA 568
Db 344 KQOQFELMDATFSSVEKVIHTTLEVIQDRI-----FGTS----- 377
QY 569 DNGDLIDLPVTEPPYASPRYLKRNFTLVTDYFSNMCAPGPHLPHTPDPPVLPSEAK 628
Db 378 -----ANPSLL-----WNSPPGA--LLALSTPS--TSPQDFP 405
QY 629 FLGSGTGFRPIGGAGSGKEFOALGQNVPEQFTVMTLYEREVLNLSLERLNGPY 688
Db 406 FYILOGGRPEG-----RSALIWGPPQGPPLKLIQAVAGSQH 444
QY 689 LNKVYVW-NSPKLPSEDLMPDIGVPIWVTRTEKNSLNKRLPWNIEIETALISIDDA 747
Db 445 CAQILVMSNERPLPSR---WPEYAVPLTVIDGHR-KVSDREFFPYSTIRTDALISIDARS 500
QY 748 HLHREIMFGFRVYREARDRIYGFPGRYHANDIPHOSWLYNSYSCELSWLTGAAPFK 807
Db 501 SLTSEVDFALVWOSPERAVGFLTSHFDEAHGKGITYAKERTNESMVLTTAAEFYHR 560
QY 808 YYAVLYSTVMPQAIRDNDVEYINCEDIANMFLVSHITRKPRIKY-----TSRWTFRCGP- 861
Db 561 YYHTLFHSHLPKALRTLADEAPTCVDVLMNFIYAAVTKLPPIKYVYKQGEAPLAPAGG 620
QY 862 -----CQOALSHDSHFERHKCINFEVKYGYMPLLYTORVSVLFK 905
Db 621 GPRPRKPAPAPD-----CINQIAAAGMPLLSRLRLDPVLFK 660

RESULT 8
EXTL_HUMAN
ID EXT1_HUMAN STANDARD; PRT; 330 AA.
AC Q9UBQ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exostosin-like 2 (EXT-related protein 2).
GN EXT12 OR EXT12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98111853; PubMed=9450183;
RA Wuyls W., Van Hul W., Hendrickx J., Speleman F., Wauters J.,
RT De Boule K., Van Roy N., Van Agtmael T., Bossuyt P., Willems P.J.;
RT "Identification and characterization of a novel member of the EXT gene
family, EXT12."
RL Eur. J. Hum. Genet. 5:382-389(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98139867; PubMed=9473480;
RA Saito T., Seki N., Yamauchi M., Tsuji S., Hayashi A., Kozuma S.,
RT Hori T.-A.;
RT "Structure, chromosome location, and expression profile of EXT1 and
EXT2, new members of the multiple exostosins gene family."
RL Biochem. Biophys. Res. Commun. 243:61-66(1998).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
reticulum (By similarity).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000416; AAC02898.1; -
DR EMBL; AB009284; BAA24081.1; -
DR MIM; 602411; -
DR InterPro; IPR004263; Exostosin.
DR Pfam; PF03016; Exostosin; 1.
KW Multigene family; Transmembrane; Signal-anchor.
FT TRANSMEM 23 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CARBOHYD 74 74 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE .330 AA; 37465 MW; 6976BE7EC6F588C8 CRC64;

Query Match 7.5%; Score 367; DB 1; Length 330;
Best Local Similarity 30.8%; Pred. No. 4.2e-19;
Matches 86; Conservative 49; Mismatches 98; Indels 46; Gaps 7;

QY 656 GNVPEQFTVMTLYEREVLNLSLERLNGLPYLNKVVYVWNSPKLPSEDLMPDIG--- 712
Db 59 GKSTMDSTLLIMQTYNRKIDLLKLNHQAAPNLAKYIVANNNGEKAPDELMNSLGPH 118
QY 713 VPINVYTRTEKNSLNKRLPWNIEIETALISIDDAHLHREIMFGFRVYREARDRIYGF 772
Db 119 IPIVFKQGTARMRRLQVPEPELNAVLDPTLISTPDLVFAFSVWQGFPPQIVGFV 178
QY 773 GRYHAMDIPHOSWLYN-----SNYSCELSWLTGAAPFKRYAYLYKSYMPQAIR 822
Db 179 PRKH---VSTSSGIYSYSGSFEMQAPSGNGDQYSWVLIGASFNSKYLELQOR-QPAVH 234
QY 823 DMVDEYINCEDIANMFLVS-HITRKPRIKY-----TSRWTFRCGPQA 865
Db 235 ALIDDTQNCDDIANMFIKAIKIGTSGIFVYPMVMDNLEKTNISGYSGM----- 284
QY 866 LSHDSHFERHKCINFEVKYGYMPLLYTORVSVLFK 904
Db 285 --HRAEHALGRSCINKLVNIDSMPLRYSNIMISQGF 321

RESULT 9
EXTL_CAEEL
ID EXT1_CAEEL STANDARD; PRT; 378 AA.
AC O01704;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Multiple exostosins homolog 1.
GN RIB-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
NC Rhabditidae; Pelodirinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97264342; PubMed=9110175;
RA Cline G.A., Ashley J.A., Shah S., Lovett M.;
RT "The structure of the human multiple exostosins 2 gene and
RT characterization of homologs in mouse and Caenorhabditis elegans."
RL Genome Res. 7:359-367(1997).
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



DE V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).  
 GN API OR AF1159.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; Pubmed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RA "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RT Nature 390:364-370(1997)  
 CC  
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON  
 CC GRADIENT ACROSS THE MEMBRANE.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
 CC H(+) (out).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE001023; AAB90076.1; -  
 DR TIGR: AF1159;  
 DR InterPro: IPR002490; V\_ATPase\_sub\_a.  
 DR Pfam: PF01496; V\_ATPase\_sub.a: 1.  
 KW Hydrolyase; Hydrogen ion transport; Transmembrane; Complete proteome.  
 FT TRANSMEM 341 361 POTENTIAL.  
 FT TRANSMEM 390 410 POTENTIAL.  
 FT TRANSMEM 449 469 POTENTIAL.  
 FT TRANSMEM 490 510 POTENTIAL.  
 FT TRANSMEM 538 558 POTENTIAL.  
 FT TRANSMEM 564 584 POTENTIAL.  
 FT TRANSMEM 590 610 POTENTIAL.  
 FT TRANSMEM 617 637 POTENTIAL.  
 SQ SEQUENCE 676 AA; 76649 MW; E2E0AC8890C9CA45 CRC64;

Query Match 2.38; Score 113; DB 1; Length 676;  
 Best Local Similarity 21.98; Prid. No. 2.3; Mismatches 124; Indels 80; Gaps 15;  
 Matches 68; Conservative 39;

OY 259 -QLY-----SLPHMRTDGHNVIIIMLSRKSPTONILY-NVSTGRAMVA---OSTFYVQYR 309  
 Db 417 YQHIFISQLEHLXEFHGHGHPFVREVMGKILLFNTFYIGAKILFGALSGYNNYVE 476  
 OY 310 PGF-DLVVSP 319  
 Db 477 HGLKDAILEKL 487  
 RESULT 12  
 CENE\_HUMAN STANDARD; PRT; 2663 AA.  
 ID CENE\_HUMAN  
 AC 002224;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Centromeric protein E (CENP-E protein).  
 GN CENPE.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93024922; Pubmed=1406971;  
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;  
 RT "CENP-E is a putative kinetochore motor that accumulates just before  
 RT mitosis.";  
 RT Nature 359:536-539(1992).  
 RL  
 RN  
 RN CHARACTERIZATION.  
 RX MEDLINE=98437347; Pubmed=9763420;  
 RA Chan G.K.T., Schaar B.T., Yen T.J.;  
 RT "Characterization of the kinetochore binding domain of CENP-E reveals  
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";  
 RL J. Cell Biol. 143:49-63(1998).  
 CC  
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
 CC AND/OR SPINDLE ELONGATION.  
 CC  
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
 CC  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z15005; CAA78727.1; -  
 DR PIR: S28261; S28261.  
 DR HSSP: P17119; 3KAR.  
 DR MIM: 117143; -  
 DR InterPro: IPR001752; kinesin.  
 DR Pfam: PF00225; kinesin; 1.  
 DR PRINTS: PR00380; KINESINHEAVY.  
 DR SMART: SM00129; KISC; 1.  
 DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

Cell cycle; Centromere.  
 KW DOMAIN 1 335 KINESIN-MOTOR.  
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
 FT NP\_BIND 86 93 ATP (BY SIMILARITY).  
 SQ SEQUENCE 2663 AA: 312087 MW: 69FC13880C86C8B CRC64:

Query Match 2.3%; Score 113; DB 1; Length 2663;  
 Best Local Similarity 18.4%; Pred. No. 18;  
 Matches 143; Conservative 114; Mismatches 308; Indels 214; Gaps 33;

54 LBDADAGKIFPPRGVNECEVKNVLDLCRIEVSSEELLQLEAKROELNSTAKLNK 113  
 1877 LHNLEEMKSVKMER--DNLRLVEETLKL--EKDQKESIQETKRAKIQEQLKATARM 1932  
 114 IECKKSIENAKODL---LQKNVSIQTEHSYKELMAQNPRLSPRLRPKDDAGLP 169  
 1933 SKHKETVQKLRKIKSEKTIQISDIOKDLKSKDELQKIQEQLKELQKELQKEDVNM 1992  
 170 PPKATGCRLANCFDYSRCPLTSGFPVYV--YDSQFVFGSYLDPVKAQATARANY 227  
 1993 HKKINEMEQKKQFE-----PVLCKCEMDNF-----QLTKLHESLEIRIV 2035  
 228 VTENAD-----IACLYVILVGENQEPVYLRPALEKQVSLPHMRDGHN 272  
 2036 AKERDELRRIKESLKMEROQFIATLRERIMARDQNHQV--KP---EKRLLS-----DGOQ 2085  
 273 HVILNLSRK--SDQNLLVNVSTGRAVAQSTFYTVQVRPGFDLVSLVHAMSEPNFMEI 331  
 2086 HLMESLREKSKRIKELKRY-----EMDDHTECLNRLSLD-----EKEIEFHRI 2131  
 332 PPOVPVKKYLTFFQGEKTESLSLQEARSEEMEG-----DPADYDDRII 380  
 2132 MKRLKTVLVSVTIKKEQHCHCKFKEMD---FIDEVEKQKELLKIQHLOQDDVPSREL 2188  
 381 ATKAVO--DSKIDQVLEVETCKNPKPSLPTW--ALCGERE-----DRLELLK 426  
 2189 RDKLQNMMDLHTEILKDFS--ESEFPSIKTEFOQVLSNRKETOFLBEWLTNRDIEK 2246  
 427 LSTFALITFGDRLVYSSGCATRLP---EALVGVAVPLVGEQVQLPYQDMQWNE-- 480  
 2247 LKNG--IQKENDNICOVNFFNNRITAIINVESTEFERSATISKEMQDLKSLKNEKL 2304  
 481 -----AALVVP-----KPRVT-----EVHFLSLSDSLAMRQG 512  
 2305 FKNVQTLKTSLSAGAQVNPITQDNKNPHVSRATQLTETKIRELENSLHAKESAMHKS 2364  
 513 RPLMETEYFSTADSIPTNTVLAMIRTRIQIPAPLREAAAEIPHRSGK----- 559  
 2365 KII---KMQELEVTNDIILKQAKV-----HESNCKLEKTKETIQVLQ 2405  
 560 ---AAGDPMAMNDGDLGPV-----ETEPYAPSPRYLNFPLTVDFY 601  
 2406 DKVALGAKPKKEEDKMKLVKIDLEKKNNAKEFEKELISATATAYEYQVIRLLRENL 2465  
 602 RSNMCAKGPFRHLPFRPFLPSEAKFLGSGTGFRPI--GGGAGSGKGFQALGQNP 659  
 2466 RRSQAOQDTSVISEHT--DP-QPSN-----KPLTGGGSGIQQNKKALIKSEHI 2512  
 660 RQPTVVMVLTREVEVLMSLERLNGLPYANKVYVW-----NSRPLP 702  
 2513 RLEKEISKLKQONQOLIKQKNELLNNQHLNSNEVKTWKERTLRKAHKQVTCENSEPSP 2571

RESULT 13  
 ID DYHC\_CAEEL STANDARD; PRT; 4568 AA.

AC 019020;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Dynein heavy chain, cytosolic (DYHC).

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=96114101; PubMed=8674131;  
 RA Lye R.J., Wilson R.K., Waterston R.H.;  
 RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the  
 RT nematode Caenorhabditis elegans."  
 RL Cell Motil. Cytoskeleton 32:26-36(1995).  
 CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
 CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
 CC ORGANELLES ALONG MICROTUBULES.  
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
 CC INTERMEDIATE AND LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: L33260; AAC37251.1; -  
 DR InterPro: IPR003593; AAA.  
 DR InterPro: IPR004273; Dynein\_heavy.  
 DR Pfam: PF03028; Dynein\_heavy; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.  
 FT DOMAIN 587 652  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 814 844  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1241 1274  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1324 1340  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1559 1591  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1964 1992  
 FT MICROTUBULE-BINDING (POTENTIAL).  
 FT DOMAIN 3132 3229  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 3339 3432  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 3707 3739  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 4359 4386  
 FT COILED COIL (POTENTIAL).  
 FT NP\_BIND 1865 1872  
 FT ATP (POTENTIAL).  
 FT NP\_BIND 2163 2170  
 FT ATP (POTENTIAL).  
 FT NP\_BIND 2537 2544  
 FT ATP (POTENTIAL).  
 FT NP\_BIND 2880 2887  
 FT ATP (POTENTIAL).  
 SQ SEQUENCE 4568 AA: 521568 MW: 028E52684F61676 CRC64;

Query Match 2.3%; Score 112.5; DB 1; Length 4568;  
 Best Local Similarity 18.7%; Pred. No. 42;  
 Matches 158; Conservative 96; Mismatches 256; Indels 335; Gaps 36;

87 ESYSEELLQLEAKROELNSTAKLNKIEACKKSIENAKODLLQKNVSIQTEHSYKELM 146  
 3353 EPLRNEIKRLQGEAAKKTQGVKVDVITELSESIGKVEEYQVL--IQAGENIKQD 3409  
 147 AQNPKLSLPIRLPPEKDAGLPKPKATRCGLHNCNDYRCPVLSFPYVVDSDQFV- 205  
 3410 SYQE-----KVNRSSTELLSLRSEDRKWSGSGAFSGOQMSVLG 3448  
 206 -----FGSYLDPVKAQ-----FQATARANYVTENADIAC--YVILVGENQEPV 249  
 3449 DALLSNFLAYAGYVQDMQDELFEHKFNHVVNAGLHFRH--DLAREVYSTVDD----- 3501  
 250 VLRPALEKQVSLPHMRDGHNVIIINLSRKSQTONLLVNVSTGRAMVAQSTFYTVQVR 309  
 3502 -----RLQWQNLNSLPVDDICTENAIMLHFRN----- 3528  
 310 PGFDLVVSPVLAHAMSEPNFMEIPQVPVKKRYLF--TFQGEKI-----ESLSSIQEAR 361

```

Db 3529 --YPLIDPSGOAV-----EYIMKQFACKNIQKTSFLDESRKULESAL 3570
QY 362 SEEMEEDPPADYD-----DRIATL-----KAVODSKLD 392
Db 3571 RENGSLVQDVEAYDPIINPVLRNREVRKAGRGVLTITQDQIDLSPPFOIFMTIDRSTVE 3630
QY 393 -----QVLEFFCKNQPKPSLPTENALGGERD-----RIELKL-STFALLITPGDP 439
Db 3631 FSPDLCRVTYNTVYVTSLSAOCLOVLSNRPDVKRNDLKLLOGEFAVRLRHLEK 3690
QY 440 RL-----VISSGCATRLFEALVGAAPV-----VLGE--QVQLPYODM-- 475
Db 3691 ALLALNESKGIIDNSVIEITLEKLNKALEVAQSAETQKMAVEVAASQYORLSTA 3750
QY 476 -----LQNNBAALVYKPRVTEVHFILRLSOSDLAMRRQGRFLMETY----- 519
Db 3751 CSHYHTLQ-----OLNEIHFLLYHSLD-----FLVEIFTHVLKTEPL 3788
QY 520 FSTAD-----SIFNTV-----LAMIRTRIOI-----PAAPITREMAA 551
Db 3789 SSTDYAKRLRLITTSLEQYVFRVRSRGMHTDKVLLMLRLHRSNPSAPAVEQHF 3848
QY 552 EIPHRSGAAGTDPNMADN-----GDLIGPVETEPYASPR-----YLRFNLTLYT 598
Db 3849 LILGRSDFVAKNDE--ADSTIPGDLFLTEVKNKSIAKARVGVGFENVAHLQHNASAVT 3906
QY 599 DEYTS-----WNCAPGPF-----HLFPHPPD--- 620
Db 3907 SWLTNDNESVYVWDADGKLSPLCIAMNSLIYVALRDRMLASAHRYVSAFAFDHF 3966
QY 621 -----PVLPSKAFKLGSGTGF-----RPIGGAGSGSGKEF- 650
Db 3967 MQQDKVVDILSYVNEVSPSEPVLLCSYTGDAAGKIEDLAVENRQULTSAISABEEN 4026
QY 651 --QAALGQNPREFYVYMLTYEREVYLMNSLERINGLP-----YLNKVVVVVNSKFL 701
Db 4027 QADALGTATKSGRMVLLKNVHLAPSWLAQLEKRLHSMKPHAQRLFLTAI-----HPKL 4082
QY 702 PSEDL 706
Db 4083 PSSIL 4087

```

```

RA Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
RA Sahnorn D., Sasaki T., Kuivanlehti H., Chu M.L., Deutzmann R.,
RA Timpl R., Uitto J.;
RT "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1
RT and B2 chains, and expression of the corresponding genes in human
RT skin and cultured cells.";
RL Lab. Invest. 60:772-782(1989).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
CC LAMININ-3 (S-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELIICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58531; CAA41418.1; -.
DR PIR: S14458; S14458.
DR HSSP: P02468; IKLO.
DR MIM: 150320; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001886; LamNT.
DR InterPro: IPR000036; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00052; Laminin_B_2.
DR Pfam: PF00053; Laminin_EGF_15.
DR Pfam: PF00054; Laminin_G_5.
DR Pfam: PF00055; Laminin_Nterm_1.
DR PRINTS: PR00011; EGF_LAMININ.
DR PRODOM: PD002082; LamNT_1.
DR PRODOM: PD003031; Laminin_B_2.
DR SMART: SM00180; EGF_Lam_14.
DR SMART: SM00001; EGF_Like_1.
DR SMART: SM00281; Lamb_2.
DR SMART: SM00282; Lamb_5.
DR SMART: SM00136; LamNT_1.
DR PROSITE: PS00022; EGF_1; 11.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01248; LAMININ_TYPE_EGF_15.
DR PROSITE: PS50025; LAM_G_DOMAIN; 5.
DR KW Laminin EGF-like domain; Cell adhesion; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 1 3075
FT DOMAIN 18 269 LAMININ ALPHA-1 CHAIN.
FT DOMAIN 270 326 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 327 396 LAMININ EGF-LIKE 1.
FT DOMAIN 397 453 LAMININ EGF-LIKE 2.
FT DOMAIN 454 502 LAMININ EGF-LIKE 3.
FT DOMAIN 503 512 LAMININ EGF-LIKE 4.
FT DOMAIN 517 708 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 709 741 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT LAMININ EGF-LIKE 5 (C-TERMINAL).

```



FT	DOMAIN	742	790	LAMININ BGF-Like 6.
FT	DOMAIN	791	848	LAMININ BGF-Like 7.
FT	DOMAIN	849	901	LAMININ BGF-Like 8.
FT	DOMAIN	902	950	LAMININ BGF-Like 9.
FT	DOMAIN	951	997	LAMININ BGF-Like 10.
FT	DOMAIN	998	1043	LAMININ BGF-Like 11.
FT	DOMAIN	1044	1089	LAMININ BGF-Like 12.
FT	DOMAIN	1090	1149	LAMININ BGF-Like 13.
FT	DOMAIN	1150	1159	LAMININ BGF-Like 14.
FT	DOMAIN	1160	1361	LAMININ BGF-Like 15.
FT	DOMAIN	1362	1402	LAMININ BGF-Like 16.
FT	DOMAIN	1403	1451	LAMININ BGF-Like 17.
FT	DOMAIN	1452	1508	LAMININ BGF-Like 18.
FT	DOMAIN	1509	1555	LAMININ BGF-Like 19.
FT	DOMAIN	1556	2116	LAMININ BGF-Like 20.
FT	DOMAIN	2117	2297	LAMININ G-Like 1.
FT	DOMAIN	2305	2481	LAMININ G-Like 2.
FT	DOMAIN	2486	2673	LAMININ G-Like 3.
FT	DOMAIN	2713	2885	LAMININ G-Like 4.
FT	DOMAIN	2890	3070	LAMININ G-Like 5.
FT	DOMAIN	1706	1796	COILED COIL (POTENTIAL).
FT	DOMAIN	1968	1989	COILED COIL (POTENTIAL).
FT	DOMAIN	2088	2120	CELL ATTACHMENT SITE.
FT	DOMAIN	2534	2536	CELL ATTACHMENT SITE.
FT	DOMAIN	270	279	BY SIMILARITY.
FT	DOMAIN	272	290	BY SIMILARITY.
FT	DOMAIN	292	301	POTENTIAL.
FT	DOMAIN	297	305	BY SIMILARITY.
FT	DOMAIN	304	324	BY SIMILARITY.
FT	DOMAIN	327	336	BY SIMILARITY.
FT	DOMAIN	329	361	BY SIMILARITY.
FT	DOMAIN	364	373	BY SIMILARITY.
FT	DOMAIN	376	394	BY SIMILARITY.
FT	DOMAIN	397	409	BY SIMILARITY.
FT	DOMAIN	399	427	BY SIMILARITY.
FT	DOMAIN	429	438	BY SIMILARITY.
FT	DOMAIN	441	451	BY SIMILARITY.
FT	DOMAIN	454	467	BY SIMILARITY.
FT	DOMAIN	471	471	BY SIMILARITY.
FT	DOMAIN	473	482	BY SIMILARITY.
FT	DOMAIN	485	500	BY SIMILARITY.
FT	DOMAIN	742	751	BY SIMILARITY.
FT	DOMAIN	744	757	BY SIMILARITY.
FT	DOMAIN	760	769	BY SIMILARITY.
FT	DOMAIN	772	788	BY SIMILARITY.
FT	DOMAIN	791	806	BY SIMILARITY.
FT	DOMAIN	793	816	BY SIMILARITY.
FT	DOMAIN	819	828	BY SIMILARITY.
FT	DOMAIN	831	846	BY SIMILARITY.
FT	DOMAIN	849	863	BY SIMILARITY.
FT	DOMAIN	851	870	BY SIMILARITY.
FT	DOMAIN	873	882	BY SIMILARITY.
FT	DOMAIN	885	899	BY SIMILARITY.
FT	DOMAIN	902	914	BY SIMILARITY.
FT	DOMAIN	904	921	BY SIMILARITY.
FT	DOMAIN	923	932	BY SIMILARITY.
FT	DOMAIN	935	948	BY SIMILARITY.
FT	DOMAIN	951	963	BY SIMILARITY.
FT	DOMAIN	953	969	BY SIMILARITY.
FT	DOMAIN	971	980	BY SIMILARITY.
FT	DOMAIN	983	995	BY SIMILARITY.
FT	DOMAIN	998	1007	BY SIMILARITY.
FT	DOMAIN	1000	1014	BY SIMILARITY.
FT	DOMAIN	1016	1025	BY SIMILARITY.
FT	DOMAIN	1028	1041	BY SIMILARITY.
FT	DOMAIN	1044	1056	BY SIMILARITY.
FT	DOMAIN	1046	1063	BY SIMILARITY.
FT	DOMAIN	1065	1074	BY SIMILARITY.
FT	DOMAIN	1077	1087	BY SIMILARITY.
FT	DOMAIN	1403	1412	BY SIMILARITY.
FT	DOMAIN	1405	1419	BY SIMILARITY.
FT	DOMAIN	1422	1431	BY SIMILARITY.
FT	DOMAIN	1434	1449	BY SIMILARITY.

FT	DISULFID	1452	1466	BY SIMILARITY.
FT	DISULFID	1454	1476	BY SIMILARITY.
FT	DISULFID	1479	1488	BY SIMILARITY.
FT	DISULFID	1491	1506	BY SIMILARITY.
FT	DISULFID	1509	1521	BY SIMILARITY.
FT	DISULFID	1511	1528	BY SIMILARITY.
FT	DISULFID	1530	1539	BY SIMILARITY.
FT	DISULFID	1542	1553	BY SIMILARITY.
FT	DISULFID	1556	1556	INTERCHAIN (PROBABLE).
FT	DISULFID	1560	1560	INTERCHAIN (PROBABLE).
FT	CARBOHYD	555	555	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	763	763	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1407	1407	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1579	1579	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1596	1596	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1678	1678	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1689	1689	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1698	1698	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1717	1717	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1804	1804	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1894	1894	N-LINKED (GLCNAC. . .)

Query Match 2.3%; Score 111.5; DB 1; Length 3075;  
 Best Local Similarity 20.0%; Pred. No. 28;  
 Matches 119; Conservative 105; Mismatches 237; Indels 135; Gaps 27;

QY	71	NECEYKVLVD-LCRRESYSEBLLQLEAKROELNSEIANKLKEACKSIENAKODLL	129
DB	1717	NATLEIKAAADLLSQIOENYOKPLEELEVLEKAAASHVLSKHNNELKAAEALVBEAEAKMQ	1776
QY	130	QKKNVI-----SOTESHYKRLMAQNPCKLSPIRLPEKODAGIPPKATRGRLHNCFD	164
DB	1777	ESNLLMLMANANREFSDKRLHVOEONLTVSELV-----OGRLDAAAOOTDAVDALE	1632
QY	185	YSRCPLTSGPPVYVYDDQFVFGS-----YLDPLVKAQAFATARANYVTEN--ADIACL	237
DB	1833	HLE-----DHQDKLLMSAKIRHHIDLVHMMSQRNVDLYRAEDHATERQRL	1861
QY	238	YVILVGMQEPVVLRAPELEKOLYSLPHMRTDG-----HNHYIINLSKRSPTQNLIN--V	291
DB	1882	ADVLVSGLEN---IRNVS LNATSAAYVHYNIOSLIESESEELARDAHRTVETSLSESIV	1938
QY	292	STGRAMVAGSTFTYVQVR-----PGFDLVVSPVHAMS--EPNFMETPO-----	334
DB	1939	SNGRAAVORSRFLKEBNNLKRKLPGTALSELURNTKTRFOENAVITROTNESSLILR	1998
QY	335	-VP-----VKRYLTFQGEKIESLRSSLOEARSFEEEMEGDPADYDRIATLK--	384
DB	1999	ALPGICDKAKTKRELAT---SASQSAVSTLRDVAAGSGLTMTAS--LSRVNTTLRET	2053
QY	385	-AVODSKLDQVLYEFTCKN-QPKPSLPTMALCGEDNRYELIKLSTPLIITPGPRL	441
DB	2054	HOLDODSTMAVTLLAGRVKVEIDQANL-----LFDRIKLPKMEENLSRNLSEIKL	2104
QY	442	VISGCAITLFEALVGVAVVLEGEVQLPYODML---ONNEALVYPRKRVTEVHLLR	498
DB	2105	LISO--ARKQASIKVA---VSADRCIRAYQPOISSTNYNTLTVLANKTQEPDNILFLYL	2159
QY	499	SLSDDLLAMR-ROGR--FLMETVYFSTADSIPTVLAMIRTRIOIPAPRIEAAAEI--	553
DB	2160	SSTASDPLAVEMRGRVAFPLMD-----LGSSTRLFEFPFPIIDNNWHSIHV	2206
QY	554	-----PHSGKAAGTDPNMAONGDLDL-----GPVEEP	582
DB	2207	AREGNIGLSVKEKMSNQKSPRTSKSPGT-ANVLVNNSTLMFVGGLGOIKKSP	2261
RESULT	15		



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:41 ; Search time 37.88 Seconds  
(without alignments)  
4196.999 Million cell updates/sec

Title: US-09-809-920-4  
Perfect score: 4873  
Sequence: 1 MTGYTLRNGAGNGGCTCM.....DSVLEKTRLPMDKTCFKFI 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: SP\_Archaea:\*  
3: SP\_Bacteria:\*  
4: SP\_Fungi:\*  
5: SP\_human:\*  
6: SP\_invertebrate:\*  
7: SP\_mammal:\*  
8: SP\_mhc:\*  
9: SP\_organelle:\*  
10: SP\_Phage:\*  
11: SP\_plant:\*  
12: SP\_rodent:\*  
13: SP\_virus:\*  
14: SP\_vertebrate:\*  
15: SP\_unclassified:\*  
16: SP\_virus:\*  
17: SP\_bacteriophage:\*  
18: SP\_bacteriophage:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4755	97.6	919	11	Q9JMA8
2	4720.5	96.9	918	11	Q9WVL6
3	2314	47.5	972	5	Q9XZ08
4	1252.5	25.7	814	5	O17920
5	1003	20.6	718	11	Q923D6
6	993	20.4	718	6	O77783
7	939	19.3	728	4	O15288
8	870.5	17.9	746	4	O9BVR9
9	870.5	17.9	746	11	Q9JKB2
10	841	17.3	717	5	O91B9
11	832	17.1	738	13	Q98SV5
12	796.5	16.3	760	5	O76796
13	795.5	16.3	760	5	O9V730
14	639	13.1	669	11	Q9JUV7
15	586	12.0	131	11	O9D7G5
16	379.5	7.8	330	11	Q9ES89

17	372.5	7.6	329	11	Q9CX90
18	360	7.4	334	10	Q9LY62
19	312.5	6.4	741	10	Q9AUK5
20	306.5	6.3	764	10	Q9L275
21	288	5.9	329	10	Q9C975
22	251.5	5.2	382	5	Q9U3J6
23	211	4.3	67	6	Q9SKP6
24	153	3.1	129	4	O00245
25	135.5	2.8	498	10	Q9FEB4
26	129	2.6	1310	10	Q949K0
27	127.5	2.6	341	10	Q94AA9
28	123.5	2.5	425	10	Q9SZK4
29	123	2.5	1568	11	Q9ESK9
30	121.5	2.5	415	10	Q94Q08
31	121.5	2.5	511	10	Q9LFL2
32	121.5	2.5	1185	4	O13135
33	121.5	2.5	1202	4	O13136
34	120.5	2.5	1257	4	O75334
35	120	2.5	460	10	O82181
36	120	2.5	893	16	O9XIG2
37	120	2.5	1267	4	O75145
38	119.5	2.5	824	13	O90778
39	119	2.4	846	4	O75130
40	118.5	2.4	432	10	O9F2J1
41	117.5	2.4	811	13	O9Y1B9
42	117	2.4	440	10	O64805
43	117	2.4	1331	17	O9YEG9
44	116	2.4	499	10	O9S565
45	114.5	2.3	1024	10	O23517

## ALIGNMENTS

RESULT	ID	Q9JMA8	PRELIMINARY:	PRT:	919 AA.
AC	Q9JMA8	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	REG RECEPTOR.				
CN	REGR.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-WISTAR; TISSUE-PANCREATIC ISLETS;				
RX	MEDLINE=20219101; PubMed=1073661;				
RA	Kobayashi S., Akiyama T., Naka K., Abe M., Tajima M., Shervani N.J.,				
RA	Ueno M., Matsuno S., Sasaki H., Takasawa S., Okamoto H.;				
RT	"Identification of a receptor for reg (regenerating gene) protein, a				
RT	pancreatic beta-cell regeneration factor."				
RL	J. Biol. Chem. 275:10723-10726(2000).				
DR	EMBL: AB033367; BAA92895.1; -				
KW	Receptor.				
SQ	SEQUENCE 919 AA; 104443 MW; 6F42EA80F5C2807 CRC64;				

Query Match 97.6%; Score 4755; DB 11; Length 919;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 894; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

Qy	1	MTGYTLRNGAGNGGCTCMRNSNRIRLTSLTFLVFPFLINHYTTTIDEADEA 60
Db	1	MTGYTLRNGAGNGGCTCMRNSNRIRLTSLTFLVFPFLINHYTTTIDEADEA 60
Oy	61	GKRFEPVGNELCEVHVLDLCRIRESVSEELQLEAKROELNSETAKLNLKTEACKS 120
Db	61	GKRFEPVGNELCEVHVLDLCRIRESVSEELQLEAKROELNSETAKLNLKTEACKS 120

```

QY 121 IENAKODLQKNTVISOEHSYKELMAONPKLSIPRLPEKDDAGLPPEKATGCRH 180
Db 121 IENAKODLQKNTVISOEHSYKELMAONPKLSIPRLPEKDDAGLPPEKATGCRH 180
QY 181 NCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPVKAQAFQATARANVYTEMADICLYVI 240
Db 181 NCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPVKAQAFQATARANVYTEMADICLYVI 240
QY 241 LVGEQOEPRVLRPALEKOLYSLPHWRTDGHNVIIINLSRKSDDONLLYNSTGRANVAQ 300
Db 241 LVGEQOEPRVLRPALEKOLYSLPHWRTDGHNVIIINLSRKSDDONLLYNSTGRANVAQ 300
QY 301 STEFTVOYRPFQDLVYSPVLAHAMSEPNMEIPPOVPKRYKLTFTFGKIESLSLOEA 360
Db 301 STEFTVOYRPFQDLVYSPVLAHAMSEPNMEIPPOVPKRYKLTFTFGKIESLSLOEA 360
QY 361 RSFEEMGDPADYDDRIITATLKAVIDSKLDQVLEFTCKNPKPSLPTMALCGERED 420
Db 361 RSFEEMGDPADYDDRIITATLKAVIDSKLDQVLEFTCKNPKPSLPTMALCGERED 420
QY 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVLEQVQPLPYOMLQWNE 480
Db 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVLEQVQPLPYOMLQWNE 480
QY 481 AALVVPKRYTEVHFILRSLSDSLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
Db 481 AALVVPKRYTEVHFILRSLSDSLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
QY 541 PAADIREBAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
Db 541 PAADIREBAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
QY 601 YRSWNCAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 660
Db 601 YRSWNCAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 660
QY 661 EGFVVMALTYEREVEVLNLSLERLNGLPYLKVVVWVNSPKLPSBDLMPDIGVIMVVRT 720
Db 661 EGFVVMALTYEREVEVLNLSLERLNGLPYLKVVVWVNSPKLPSBDLMPDIGVIMVVRT 720
QY 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWRREARDRIYVGPGRYHAMD 780
Db 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWRREARDRIYVGPGRYHAMD 780
QY 781 PHOSWLYNSVSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
Db 781 PHOSWLYNSVSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
QY 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDHHERHKCINFEVYKYGWMLLYTOFRVD 900
Db 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDHHERHKCINFEVYKYGWMLLYTOFRVD 900
QY 901 SVLEKTRLPDKTKCFKFI 919
Db 901 SVLEKTRLPDKTKCFKFI 919

```

```

RT "Molecular cloning of the mouse homolog of ExTL1."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083550; AAD42040.1; -.
DR MGI: 1860765; ExTL1.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
SQ SEQUENCE 918 AA; 104547 MW; 4EEFEC86E7E6027E CRC64;

Query Match          96.9%; Score 4720.5; DB 11; Length 918;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 891; Conservative 7; Mismatches 20; Indels 1; Gaps 1;

QY 1 MTGYTMLRNGAGNGGQCMRLWNSRIRLTMLSFTLVLFVFFPLAHYLLTIDEADEA 60
Db 1 MTGYTMLRNGAGNGGQCMRLWNSRIRLTMLSFTLVLFVFFPLAHYLLTIDEADEA 60
QY 61 GKRIEPRVAGELCEVKVAVDLCRIRESVSEBLLQLEAKROELNSETAKLMKIEACKS 120
Db 61 GKRIEPRVAGELCEVKVAVDLCRIRESVSEBLLQLEAKROELNSETAKLMKIEACKS 120
QY 121 IENAKODLQKNTVISOEHSYKELMAONPKLSIPRLPEKDDAGLPPEKATGCRH 180
Db 121 IENAKODLQKNTVISOEHSYKELMAONPKLSIPRLPEKDDAGLPPEKATGCRH 180
QY 181 NCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPVKAQAFQATARANVYTEMADICLYVI 240
Db 181 NCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPVKAQAFQATARANVYTEMADICLYVI 240
QY 241 LVGEQOEPRVLRPALEKOLYSLPHWRTDGHNVIIINLSRKSDDONLLYNSTGRANVAQ 300
Db 241 LVGEQOEPRVLRPALEKOLYSLPHWRTDGHNVIIINLSRKSDDONLLYNSTGRANVAQ 300
QY 301 STEFTVOYRPFQDLVYSPVLAHAMSEPNMEIPPOVPKRYKLTFTFGKIESLSLOEA 360
Db 301 STEFTVOYRPFQDLVYSPVLAHAMSEPNMEIPPOVPKRYKLTFTFGKIESLSLOEA 360
QY 360 RSFEEMGDPADYDDRIITATLKAVIDSKLDQVLEFTCKNPKPSLPTMALCGERED 419
Db 360 RSFEEMGDPADYDDRIITATLKAVIDSKLDQVLEFTCKNPKPSLPTMALCGERED 419
QY 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVLEQVQPLPYOMLQWNE 480
Db 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVLEQVQPLPYOMLQWNE 480
QY 481 AALVVPKRYTEVHFILRSLSDSLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
Db 481 AALVVPKRYTEVHFILRSLSDSLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
QY 541 PAADIREBAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
Db 541 PAADIREBAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
QY 601 YRSWNCAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 660
Db 601 YRSWNCAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 660
QY 660 YRGWNSAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 659
Db 660 YRGWNSAPGPHLPHTPPDPVLPSEAKFLGSGTGFRRIGGAGSGSKEFOALGAVNR 659
QY 661 EGFVVMALTYEREVEVLNLSLERLNGLPYLKVVVWVNSPKLPSBDLMPDIGVIMVVRT 720
Db 661 EGFVVMALTYEREVEVLNLSLERLNGLPYLKVVVWVNSPKLPSBDLMPDIGVIMVVRT 720
QY 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWRREARDRIYVGPGRYHAMD 780
Db 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWRREARDRIYVGPGRYHAMD 780
QY 781 PHOSWLYNSVSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
Db 781 PHOSWLYNSVSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
QY 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDHHERHKCINFEVYKYGWMLLYTOFRVD 900
Db 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDHHERHKCINFEVYKYGWMLLYTOFRVD 900

```

RESULT 2  
 Q9WVL6 PRELIMINARY; PRT: 918 AA.  
 AC Q9WVL6;  
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
 DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE EXOSTOSIN EXT1.  
 GN EXTL1.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10960;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Sato T.;  
 Applicant

QY 901 SVLEKTRLPDKTKCKEFT 919  
Db 900 SVLEKTRLPDKTKCKEFT 918

RESULT 3  
ID 09X208 PRELIMINARY; PRT; 972 AA.  
AC 09X208;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE BCDNA:LD21192 OR CG15110.  
GN BOTV OR BCDNA:LD21192 OR CG15110.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephyroidae; Drosophilidae; Drosophila.  
NCBI\_TextId=7227;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; Pubmed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Arroll J.F., Ashby A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Boulter J., Brokstein P., Brotler P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Donnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertilla S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Haritz N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
RA Jatalin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen M.G.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
[2]  
SEQUENCE FROM N.A.  
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
RA Abmayyan A., Arcalata T.T., Baxter E., Blazer R.G., Butenhoff C.,  
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,  
RA Galle R.A., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,  
RA Park S., Segueira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,  
RA Celniker S.E.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: A0003797; AAF57601.1; -;  
DR EMBL: AF132161; AAD34749.1; -;

DR FlyBase: FBgn0027535; botv.  
DR Interpro: IPR004263; Exostosin.  
DR Pfam: PF03016; Exostosin; 1.  
DR Hypothetical protein.  
SQ SEQUENCE 972 AA; 110999 MW; 2A11BD6912FFCA CRC64;

Query Match 47.5%; Score 2314; DB 5; Length 972;  
Best local similarity 48.1%; Pred. No. 3.5e-177;  
Matches 473; Conservative 153; Mismatches 252; Indels 106; Gaps 21;

QY 11 GAGGCGTC-----MLRKNRIRLW-----LSFLVLYVFPFLIAHYLTLD 55  
Db 20 GGGGNGACAPNSSSAQIRSMGFRSMROFRYKLPWVLMFLVSLCAIRILVEQ 79  
QY 56 EAD-----EAGKRFGRVGNELCEVHVLDLCIRISSEELLQLEAKROELN 104  
Db 80 DAPPLDLHRSSPLLDAYEDFSAMRAGDLKRIE---EMVRIKTSVSVLENELESRRQK 136  
QY 105 SETAKNLKTIACKSKSIENAKODLLQKNVISQTESHYKELMAONKLSLPIRLPEKD 164  
Db 137 SDISQYNQKIEELKQELRLREOTELERLIKISVEQAOVAREAVORNTPDALPRSLPNT- 195  
QY 165 DAGLPPK-----ATRGRLHNGEDYSRCLTSGFPVYVYDSQDFV---GSYLDPLVK 215  
Db 196 ---LPRKSNPTGMAASCEHNGFNHSCSLSGFPVYLYLDPDEHSVORKYDIDGFLK 252  
QY 216 QAFQATARAVYVENADIACLVYLVE--MQEPVLR----- 252  
Db 253 TTKLQKQTVGNVHYKDKRHCICYLVGELLEDLLRNRYAQAQEAHQPSPTLEND 312  
QY 253 -PALEKQVLSLPRKRDGNHVIINLSK---SDTONLYNVSTGMAAOSTFTYVQY 308  
Db 313 CPVMEK-LVSLPYMGDGRNHYLLNLRDLTSHRNPLRYQMTMAIVQSAFEERQF 371  
QY 309 RPFGLVSPVYVHMSPNFMEIPQVYVKKRYLFTQGEKIESLSLQARFEEEME 368  
Db 372 RFGDLVLPPLGPGGDWQECAMWPAKKYLLTQGE-LRKQSLNPLDAFILEHL 430  
QY 369 GDEPADYDRIATLKAVDKSLDOVLVEFTG---KNQPKSLPTMALCGEREDLET 424  
Db 431 AD-----MAKATQDQFVLQFCVPAEQEDSDLP-DWTLGSGSSSRRL 475  
QY 425 LKSLTFLIITPGDPRVYISGCFRLEALEVCAVYVVG-EDVQLPYQDMQMNAL 483  
Db 476 LKDSFTLILPLNGR-VSSFLMARIVYEAIRSAVAVPICADLRVLYMETVMRRTAL 534  
QY 484 VVPRVTEVHFLRSLSDSLAMRQGRFLMFTYFSTDSINFTVLAUVRTRIOVPA 543  
Db 535 LTPARITELHFLRAVQADADLLRKQGLIMERTYSSVOATVYIASLRDLGTPPR 594  
QY 544 PIREAAAE-----IPRSGKAAGTDPNMDNGDLGPEVTEPPYASPRYLRNFTLV 597  
Db 595 PVPEVIAQSVNSYFLLKSDPYGLDTEPEES---LGPI--BPVPSPAFRRNYTLR 648  
QY 598 TDFPRSNACAPRPHLPHPFEDVLPSEAKFLCGSGFRPRTGAGGSCKEFQAALGN 657  
Db 649 MQAKEANNMDLPPYLYLPOLPFDLPALPSEAKFMGSHGFRIGLGLGAGKEFEESLGN 708  
QY 658 VPRQFTVVMVLYREVYVINSLSRLNGLPYLVNVMVWNSPKLPSBLLMPDVGVMV 717  
Db 709 YPRQFTVVMVLYREVYVINSLSRLNGLPYLVNVMVWNSPKLPSBLLMPDVGVMV 768  
QY 718 VRTKNSLNNFLPWNIEETALISIDDAHLRDEIMFGFRVREARDAIVGPGRYHA 777  
Db 769 LPARNSLNNRFLPFDVIEFPAVSVDDAHLRDELDFGRVVRERDRVYRPGRYHA 828  
QY 778 WDI--PHQSWLYNSNYCELSVLTGAAPFHKKYAYLYSVYPOALIDMVDYENCDIA 835  
Db 829 WDLQNPNGOMHYNSNYCELSVLTGAAPFHKKYAYLYTYTHLPQAIKDYKDYENCDIA 888  
QY 836 MNEFVSHITRPRIKYVSRMTFRGCGPOLSHDSHPHEHNRKINFEVYVGYGMPLTYT 895

Dd	889	MNFI <del>VSHITR</del> KRPVKVTSRWTFRRPGCPVLSEDDTHFQERHNCINFSRVFGTPLLNT	948
Qy	896	OEPDVSYLEFKTRLPDKTKCKET	919
		:      :       :       :	
Dd	949	QYRADSILEKTRIPHDKKCKEPT	972
RESULT	4		
017920			
ID	017920	PRELIMINARY;	PRT; 814 AA.
AC	017920;		
Dt	01-Jan-1998	(TREMBLrel. 05, Created)	
Dt	01-Jan-1998	(TREMBLrel. 05, Last sequence update)	
Dt	01-Dec-2001	(TREMBLrel. 19, Last annotation update)	
DE	K01G5_6	PROTEIN.	
GN	K01G5_6		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea;	Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloiderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		

RL Submitted (MAR-1997) to the EMBL/GenBank/DDBB databases.  
RN  
RN [2]  
RP SEQUENCE FROM N.A.  
RP MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology."; Science 282:2012-2018(1998).  
RL Science 282:2012-2018(1998).  
DR EMBL: Z92803; CABOT245.1.-.  
DR InterPro: IPR004263; Exostosin.  
DR Pfam: PF03016; Exostosin\_1.  
SQ  
SQ SEQUENCE 814 AA; 94196 MW; 38FA5BC5EB17DB16 CRC64;

Query Match	25.7%	Score 1252.5;	DB 5;	Length 814;
Best Local Similarity	33.4%;	Pred. No. 7.8e-92;		
Matches 313; Conservative 147;		Mismatches 312;	Indels 165;	Gaps 25;

QY	21	L RMSNRIRLWLSPTLLVILVEFPL	IAHYLL-----	TLDEADEACKRI	64
		:   :   :   :   :   :   :	:   :   :   :   :   :   :	:   :   :   :   :   :   :	
Db	5	LNGSSRFVSLRKSAFLIFVFIYITII	YNNSSPESWITODALKOMENTDYDASC	64	
QY	65	FGPRVGNELCEVKAHVDLCRIRESVSE	ELIOTAEAKROENSEFAKINLKEACKSIEN	123	
Db	65	SGYSIGIRLREOKRIILASVRLETESQVK	IEEIRFTQVEOIRLIRKQNELSALEGEIEA	124	
QY	124	AKODILQKWNVISOTESHSEYKELMANQ	PKISLPIRLPEKDAGLPPPKATRCGRHNCF	18	
Db	125	AQRQL-----	ELNETQWVKYKFLPRS-PIQIPRELEPQS	166	
QY	184	DYSCRPLTSGFPYUYVDSDQFVFGSYLD	PLVKAFQATARANYVTENADIACYLIVYG	243	
Db	169	DYSCRSSISSMPYVY-----		183	
QY	244	EMQEPVYLRAELAEKOLSLPHMRTGSHN	YVITNLKSSTONILYVSTGRAMVAOSTF	303	
Db	184	-----DLITSGOSEKELNV-----	FOEYIPNLVETPPKACIKIHNGIAS-PRWTF	230	
QY	304	YTVQYRGRGFDLVYSPLVHAMSEPFMEI	RPQVYVKRKLYTFQGEKIESLRSSHOEARSF	363	
Db	231	NSILFNVG-----SPLIN-----	FOSKSIHQASKI--RSF	259	
QY	364	EEMEGDPPADYD-----DRI-IATLKA	VODSKLDQVLVEFTCKN-----QPKPS-LPT	410	
Db	260	-----DFPDVNHIAVEKVDLTPRLP	FORENLISLIVDTEINFSAFSSLSAEPSSRP	313	
QY	411	EMALGGER-----EDRLLELKLSTF	ALLITPGDRIVYISSCARTLEALEVGA	466	
Db	314	VIYKSCQENCSLERRKQLIGSSFTFCFLP	-----SEMFODTLSILOGLIPILSN	365	

```

OY 467 OYUUEYOMLQWNEALVYPPRBYVEVHFLRLSLSDSLAMRQGRFLWETFEYSTDST 526
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 366 SOLLEFODLIDRRRTYRLPLARPEAHF IVQSEJIDIIEMRRVGRFLFETYLAHRHL 425
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 527 FNTVAMIRTRIOIPAAEIREAAAEIPIHRSKA -AGTDPMDANGOLD -LGVETERP 583
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 426 ASLLAALRYKQIPIREVRNRQAIPLFNSSFTAPKSVYVQVGNPDDEVILLER 483
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 584 YASPRYLNFLITYDFRKNWCAAGPHLEPHHTPEDPVLPSEAKFL -GSGTGERPIGG 642
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 484 VESTIYAVNEFEFOLYSDFENNIIMSPHYKEFLVNAAEIPLTEAEFFEDRYKGFRRP --- 540
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 643 AGGSKEFOALGVNPREOFTYVMYLYREREYVANSJERNGJAPYKNTKVVVWNSKLP 702
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 541 EPGSGAEFSKALGCRQROEQFTYVLLITERKAVLTGLALERHOLPYLANKIIVYVWNNVNRD 600
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 703 SEDLLMPDGYVIMVYRTEKNSLNNRPLMNEIETEAIIISIDDAHLHDEIMFGRVWR 762
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 601 PPD -TWSPLEHPIVEIRVAENNNLRNREVPMDRIETEAVALSIDDIDLMQOEIILAFRVR 659
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 763 EARDRIVEFQRYAHAMDIPHQSMLYNSYSCSEIWMVLTGAAPFKKYAYLYVVMQAIR 822
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 660 ENRDRIVEFPARHHA --RYGDSMFTNSNHTQOMSLTGAFFHKNTLNTAYTEMPEIR 717
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 823 DMYDEYINCEDIAMNVLVSHTRKPPIKYTSRWTFRCPGCQALSHDSHFHRRACINF 882
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 718 EHVNSIKRCEDIAMNYLVSHTRKPPIKTISRWMLKCPCTSESLEYEGTFEKRHDCMRL 777
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 883 FVKYGYMPLLYTOFPRDVSYLFRKRLRHDHDKCEKFT 919
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 778 FTKIYGNPLKFSQFADSLIEFKRLPONNOKCEKYY 814
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT	5		
0923D6			
ID	0923D6	PRELIMINARY;	PRT; 718 AA.
AC	0923D6.		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	EXOSTOSES (MULTIPLE) 2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R.;		
RL	Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.		
DQ	EMBL; BC006557; AA06557.1; -		
SEQUENCE	718 AA; 82034 MW; 58722A09BD159CFB CRC64;		

	Query Match	20.6%	Score 1003:	DB 11:	Length 718:	
	Best Local Similarity	33.2%	Pred. No. 7.8e-72:			
	Matches 260:	Conservative 125:	Mismatches 255:	Indels 168:	Gaps 27:	
QY	133 NVISQTEHSYKELMAQNPKLSLP-I-RLLPEKDDAGLPPKATRCGRCLNCFDYSRCPYL	191	:	:	:	:
Db	49 SISSSDGVEKRKSIREVPVRLPFPDSIPERGD-----SCRMHCEDEVYRC---	96	:	:	:	:
QY	192 SGF-----PYVVYDSDFV--FGSYLDPLPVKAQAQTARA---NVYYTENMDIACLYVI	240	:	:	:	:
Db	97 -GFNPKNKIKEYITPLKRYVDADGAPVSASLRREYNELLTALSDDSDDYTDDINRACTLEP	155	:	:	:	:
QY	241 LVGBM-OEPVVLRAPELEKOLSYLPHWRITDCHNNHIYIN-LSRKSDPTQLMLYVSTGRAMV	298	:	:	:	:
Db	156 SIDVLNQNP--LRKETPAQALAQLSRW-DRGTHMLEFMLEGAPPDYMTALDPDRALL	212	:	:	:	:
QY	299 AQSFTFYVQVRPGDL---VSPLVHANSEPNEMEIIPQVPRKKRYLTFEGEKIESLRS	355	:	:	:	:
	213 AGGGFSWTVTRQGADVSLPNSPLSAEWA-----LPEKAGGPRIYFL-----LSS	257	:	:	:	:



```

OY 356 SLOEARSSEEMEDPADYODDRITATLKAVODSKIDOVUVERTCKNOKRPSIPTMALC 415
Db 258 OMATHPEVREBLE-----ALQAKHOESVLTIDKCTNLSBGVLSVR-KRC 300
OY 416 GERE--DRELLKLTSTALITTPGDRPLVITSSCATRLEFALEVGAVPVVLGEQVLPYQ 473
Db 301 HOHQHFDPYDQAEATCTVLVR--GARTL-----GQAV--LSVDLQAGCVPAIDSYLTIPS 353
OY 474 DMLQWNEALVVRPRVTEVHEHFLRLSDSDLLAMRGQFHEMETEFSRADSIFNVVLAM 533
Db 354 EYLDKRRASVAVPEEKMSDVSTIDNTIPORQIETEMORARWEAEAFQSITKALATQDI 413
OY 534 IRTRIQIPAAPIREEAALAEIPIHRSKGAACTDPMADNGDLDLGRVETEPBPVAPSRYLNF 593
Db 414 INDRY-----PYAALS----- 425
OY 594 TLTVTDFPRSNACAPRPHLPHTPR-FDPLVPSBEAKFLGSGTGFRPIGGAGSGKEFOA 652
Db 426 -----YEEENDPRAKMAVSNSNPLRLPIIPQS----- 453
OY 653 ALGQVNPREEQTVVMLVYEREVEVLMNSLEKPLNGPLVKNVVMVNSP-KLPSDILLMDPI 711
Db 454 -----QGTALVLTLYDRESLEFRITEYESKVPSPSLKLVVMNNKNNPPEESIMPTI 505
OY 712 GVPILVYVETKNSLNNRPLPWNIEIETEAISIDD-AHLRHDEIMEGFRVYRREARDRIWG 770
Db 506 RVLPLVVYVATENKLSNRRFPYDEIETEAVALAIDDDIIMLTSDLOPGCYEVMRREFPDRLVG 565
OY 771 FGRGRHMDIRHQSMKLVNSNSCGLSVYLGAPFHKYVYLYLSYVMPQALNDMDYEYN 830
Db 566 YGRRLHMDHEMKMKKESEETNSVSVLLGAPFHNKYEYVLTVMMPDCKIKWDAHKN 625
OY 831 CEDIANMLVSHITRKPIKVTSHMFRCPGCE--QALSHDSDHFHRRKCKINFVKKVG 888
Db 626 CEDIANMLVAVNGKAVIKVTPRKPKCKCEPCALIDGLSDQIHMYERSGCINKKFAVSVG 685
OY 889 YMPLLTYOFRVDSVLEKTRLPDHTKCF 916
Db 686 TMLPLVVEHRADPVLVKDDPF-EKLKSF 712

RESULT 6
077783 PRELIMINARY: PRT: 718 AA.
AC 077783:
DT 01-NOV-1998 (TREMBLrel, 08, Created)
RT 01-NOV-1998 (TREMBLrel, 08, Last sequence update)
DL 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE GLUCURONYL/N-ACETYLGLUCOSAMINYL TRANSFERASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleia; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
OX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Lind T., Tulero F., McCormick C., Lindahl U., Lidholt K.;
RT "The putative tumor suppressor Ex1 and Ex2 are glycosyltransferases
RT required for the biosynthesis of heparan sulfate.";
RL J. Biol. Chem. 0:0-0(1998).
DR EMBL: AF089748; AAC5386.1; -.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
KW Transferase.
SQ SEQUENCE 718 AA; 8186 MW; D6C18AC9C7AAD971 CRC64;

```

Query Match 20.4%; Score 993; DB 6; Length 718;  
 Best Local Similarity 32.1%; Pred. NO. 5e-71;  
 Matches 258; Conservative 130; Mismatches 257; Indels 158; Gaps 28.

```

Db      49 SISSGDMVSVEKRTGRDVP/LR/LR/PADSPYBERGL-----SCMHHCFCFVYHC--- 96
Qy      192 SGF-----PVYVYDSQEV--FGSYLDPLVKQAFQATARA---NYVTEMADIACLYVT 240
Db      97 -GFNFKNKIKYIYPLKKYVGEAGVPVSSJTISREYNELTAISDSYDDYTDV/RACLFVY 155
Qy      241 LVGEQGEVVL/RP/ALEKQ/LSLPHKMTDGNHNYIIN-LSRKSPTOULLYNVSTGRAMVA 299
Db      156 SI-DLLNONS/LR/VK/ETQAOLQAQLSRW--DRGNNHLLFNMKLGCPDPDYATALDVRDRLLA 213
Qy      300 OSTEFTVQYRGEPDULVYSP/LHMASEPNFMPEIPQOVVKKRYLFTFGCKEIESLRSLQE 359
Db      214 GGGFTWYTRGQYGVST-PLYSP/LSAE--VDLPKGGKPPRYEL-----LSSOVAL 261
Qy      360 ARSFEENEGDPADYDDRIAT/LAKAVODSKLDQVLEFTCKNQPKPSLPTENALGGERE 419
Db      262 HPEYED-----LALQARIGEA/VL/LDKCSNLS-E-GVPAARRCHQOQ 304
Qy      420 --DRLELLKSLT/FALITPGDPR/LYISGCATRLFE/LGVANP/VY/GEVOQLPYDMLO 477
Db      305 AFDYPOV/LQ/ETFCMVLR--GARL-----GQAV--LS/DVLRAGCVPIIADSYVL/PFSEVD 357
Qy      478 MNEALVY/PK/RVNEVH/LRLS/DSOLLMRROGR/LMTEYESTADIS/NTV/LAM/RT/R 537
Db      358 WKRAVVVYPERKMSDYSVSILOSDRROI/EEMQROAR/MWEATYFQSIKALAT/LQIINDR 417
Qy      538 IOI/PAP/IREBAAAEI/PHRSKAGACTOPNMNANDL/DLGVETEP/YPASPRY/LRNF/LTV 597
Db      418 I-Y/P/ALSYE-----DMN-----DPAV----- 435
Qy      598 TDEYRSNMCADG/P/HL/P/HP/PD/PDLSEAK/FLGSGT/GF/RIGGAGGSGKEFOALG/N 657
Db      436 ----KMGVS/NP/LFL-----PLIPQOS----- 453
Qy      658 VP/REOFTVYV/LT/EE/EV/LANS/LER/LNG/LP/LNK/VVVVYVWNSP-KLPS/ED/LMP/DIGYPI/M 716
Db      454 ---OQFT/VILV/LTYR/VES/LF/VIT/EVS/KVPS/LK/LV/VVWNNQNNK/NP/EDSLMP/IRVPLK 510
Qy      717 VYRTEKNS/LN/RP/LM/NNEIT/EATLSIDDD-AH/LRHEIE/MGFVWMEAR/DRIY/PFGRY 775
Db      511 VYRTEKNS/LN/RP/PY/DIET/EAV/LAIDDD/LIMLT/SELO/RGYEVM/RFP/DR/LGYGR/L 570
Qy      776 HAMDIP/HOS/WL/NSY/SC/LS/MV/LTGA/FFHK/YU/LYSYVMPQAL/NDV/DEX/INEDIA 835
Db      571 HLMDEEM/KMKY/ES/EMT/EVS/MV/LTGA/AFYK/YEN/YLT/YTKMP/ED/KNW/D/HA/NHNC/DIA 630
Qy      836 MNFLVSH/TKR/P/LIKY/MS/KMT/FC/PGCP--QALSHDS/HFHR/KKCL/NF/VK/YGYWPL 893
Db      631 MNFLVANTYGA/VIKY/P/RK/KF/KC/PECT/ALIDG/LSLDOT/HVER/SEC/INK/FASV/GT/MPLK 690
Qy      894 YTOF/RVSD/LFK/TRL/RPH/KTKCF 916
Db      691 VYERAD/V/LTKD/PF-EKLS/EF 712

```

RESULT	7		
ID	015288	PRELIMINARY;	PRT; 728 AA.
AC	015288;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	MULTIPLE EXOSOMES TYPE II PROTEIN EXT2.1.		
OS	Homo sapiens (human) .		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidi; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Deng H.-X., Fan C., Xia J., Xu L., He X., Ruan Q.;		
RT	"An isoform of a candidate gene for multiple exosomoses type II.";		
RL	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.		



Db 720 SQMRDLPVLEFKDOY 733

RESULT 9

ID 09JK82 PRELIMINARY: PRT: 746 AA.

AC 09JK82;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE EXOSTOSIN 1.

GN EXTL.

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Cricetus.

NCBI\_TaxID=10029;

NCBI [1]

SEQUENCE FROM N.A.

Mei G., Bai X., Esko J.D.,

"Expression Cloning of Heparan Sulfate Copolymerase (EXT1) and Analysis of Mutations in Chinese Hamster Ovary Cell Mutants Defective in Heparan Sulfate Biosynthesis."

RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL: AF252858; AAF71276.1; -

DR InterPro: IPR004263; Exostosin.

DR Pfam: PFO3016; Exostosin 1.

SQ SEQUENCE 746 AA; 86188 MW; C3697BA4A21DA4F2 CRC64;

Query Match 17.98; Score 870.5; DB 11; Length 746;

Best Local Similarity 28.88; Pired. No. 3.9e-61;

Matches 233; Conservative 141; Mismatches 231; Indels 201; Gaps 31;

QY 162 EKDDAGL-PPPKATRG-----CRHNCEDYSCPLTSGFPVYV--DSQPFVGS 208

Db 72 ENDSGVHVSPPKRDANSSVYKCKKCMSCFPALCK-KNGKVVYVPOQKGEKTAES 130

QY 209 YLDPVLCQAFQATRAVYVYENADIAVLVIGEMQEPVLRPA--ELEKOLYSLPH 265

Db 131 Y-----QNLIAIEGSFETSDPSQACFLVLSLDTLDRD-QLEPQVYVHNRSAVQSHL 183

QY 266 WRGCGHNVITL--SRKSTQVLYLVSTGRMAVASTTYVQYRPGFLVVSPLY-- 320

Db 184 W-NNGRNLHFLNLSGTPWPTDTEVDGFI--GQAMLAKASISTENFRNPFVSI-PLFSK 239

QY 321 ---HAMSPNEIPRQPVYRKLYFTFOGEKI-----ESLRSSIQEARSPSEEGEDPP 372

Db 240 DHPRTGGERGLK-NTIPRLAKMLVFKKRYLTGIGSDPRMLVHYHNGEDVL----- 293

QY 373 ADYDRIATLTK-----AVODSKLDQVLEFETCKNKPSPSLPTFWALCGEREDRLLELK 426

Db 294 -----LLTTCKHGKMDQKHKSDRCRDNTREY-----EKYDREMLH 329

QY 427 LSTFALLITPDDPRVLVSSGATRLFEALVEGANPVVLGBOVOLPYODMLQWNEALVVP 486

Db 330 NATCLV-PPGRRL-----GSFRFLALQAAVPMVLSNCGMELPFSEVIMWNOAAVIGD 382

QY 487 KPRTEVHFLRLSDSLDLMARQGRFLMETYSTADSFNTVLAIRTRIQIPAPIR 546

Db 383 ERLQLQIPSTRSHQKILALROTOFLMAYFSSVEKYVLTLEIIOQDI----- 434

QY 547 EEAALIEPHRSKAAGTDPNMADGDLGLGFEVTEPPYASPRYLNFLLTVTDFEYSNMC 606

Db 435 -----FKHIS-----RNSLI-----WNK 447

QY 607 ACPGPHLPHPDPVLPSEAKPIG-----SGTGFRPIGGAGSGSGEPALAGNVP 660

Db 448 HPG-----GLF-----VLPQYSSYLGDPEYYVYANGLRP-----PSKFAVAVIHAVTP- 489

QY 661 EQFVVMVLTREVLVMSLERLNGLPYLNKVVVWVNSPK-LPSEDLIMPDIQPIVIVR 719

Db 490 -----LVSQSQPVKLVLVAAAKSQYCAQIYVLMNCDKRLPAKH-RWPAIVAVYVIE 540

QY 720 TEKNSLNRRLEPWNIEETAILSIDDAHLRHDEIMGFRVWRREARDRIYGFPGRYHAM 779

Db 541 GESKVMSSRFLPYNITTDVAVLSDEDTVLTSTVEVFAFYTWQSFPRYIGYPARSHFMD 600

QY 780 IPHOSWLYNSNSYCELSMVLTGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFL 839

Db 601 NSKERMGVTSKMTNDYSVMVLGAIVHKKYHYLYTHVLPASLKNMVQQLNCEDIINFL 660

QY 840 VSHTRKPKIKVT-----SNWTRFCGCGQALSHDSSHFERHKCINFP 883

Db 661 VSAVTKLPPIKVTQKKQYKETMVGQTSRASHA-----DDHFAQRQSCMNTF 708

QY 884 VKVGYMPLLYTOFRVDSVLEKTRL 908

Db 709 ASWEGVMPLIHSQMRDLPVLEFKDOY 733

RESULT 10

ID 09Y169 PRELIMINARY: PRT: 717 AA.

AC 09Y169;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE BCDNA:GH02288 PROTEIN.

GN BCDNA:GH02288 OR CG8433.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydrioidae; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

NCBI [1]

SEQUENCE FROM N.A.

RP STRAIN=BERKELEY.

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,

RA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abell J.F., Agbayani A., An H.-U., Andrews-Plamkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokslein P., Brothier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Golder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Iregyan C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Markov G., Malshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen T.,

RA Shue B.C., Siden-Klamos I., Simpson M.C., Skupski M.P., Smith T.,

RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,

RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.N., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).  
RP SEQUENCE FROM N.A.  
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
RA Agabayan A., Arcaina T.T., Baxter E., Blazek R.G., Bultenoff C.,  
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,  
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,  
RA Park S., Sequelita A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,  
RA Celinkner S.E.;  
RT "Full length Drosophila melanogaster cDNA sequence."  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF003808; AAF58056.1; -  
DR EMBL: AF145598; AAD38573.1; -  
DR FLYBase: FBgn0027613; BCNDA:GH02288.  
DR InterPro: IPR004263; Exostosin.  
DR Pfam: PF03016; Exostosin; 1.  
SQ SEQUENCE 717 AA; 82725 MW; A3894493C16EF7DC CRC64;

## Query Match

Best Local Similarity 17.3%; Score 841; DB 5; Length 717;  
Matches 227; Conservative 126; Mismatches 238; Indels 184; Gaps 24;

QY 177 CRLHNCDFYSCPLTSGEPVYVDSDFV-----FGSYLDPLVKAQAFQATARAANYVTE 230  
DB 92 CTFWDCLNTYKCE-HDRLKYIYPLQGFVDSQDKTATTLSEYFQJLEAVLKS-R-YTTS 149  
QY 231 NADYACLYVILVGEVMOEPVYLRPAELEKQ-----LYSLPHRTGHHNVIIN-LSRKSDF 284  
DB 150 NPNACFLPISLIDLNNQV-----FDKHLGALALASIDFW-DIGANHIITNMLPGCAPS 202  
QY 285 QNLLYVSTGRANVAOSTFTYVQYRPGFDLVY---SPLY---HANSEPNMEIRPPQV 336  
DB 203 YNVLVDVNTDAIIFGGGFSWSYRPGFDVAIPWSPRLVQAHAAQAKRELLVVAQLN 262  
QY 337 YKRYLYTFPGCEKIESLRSSIQEAKSFEEEMEGDPADYDRIITATLKAVODSKLDQVLY 396  
DB 263 IL-----PRFVRTRELSLAHSEQLL 284  
QY 397 EFTCKNPKPSLPTEMALCGERDRLE---LLKLTFTALI---TTPDPRLVISGCATR 450  
DB 285 LGACENIDL---TMRCPISQHHKSLFVRLSRGKFCILGRSLRMQOPDY----- 332  
QY 451 LFEALVGAIVPVVLEGEVOLPYODMLQWNEALVVPKRYTEVHLLRSISDSDDLAMRR 510  
DB 333 --ELMSQHCIPVIVADVVPFEDVIDMSLASVRIRENEHLSVMQKLAISVXIVEMOK 390  
QY 511 QGRFLMETYESTADSIINTYIAMIRIQIPAPRIREAAAEIHRSGKAAGTDPNNADN 570  
DB 391 QVQWLESKYRKDKTLYLTALVELESRI---PFLRARS----- 426  
QY 571 GDLDLGEVTEPPYASPRYRNFTLVTDVFRSNCAPGFHLPHTPPDPV-LPSEAKF 629  
DB 427 -----RQMN---IIDNRARSTFNPFLPISLA-- 449  
QY 630 LSGGTGRPIIGGAGSGKEFOALAGNVPREQ-FTVVMLTREEREVLNMSLERLNGLPY 688  
DB 450 -----PKSGGFTAVILTYDREVESLFLLIQILAVPS 480  
QY 689 LNKVVVVVWNS-PKLPSDDLMPDIGVPIVYVTRTEKNSLNRLPWNIEFEALISIDD- 746  
DB 481 LOSILVIMNNOKSPHLLSTFPPSISKPLKIRQTKENKLSNRFYPPYPIETFEALITLIDDI 540  
QY 747 AHRHDEIMNGFWRVREARIRIVGPGRYHAMIDPHQSMLYNSYSGESEMVLGAAFFH 806  
DB 541 IMTLTDELDFGEYWRVREPHIVGFSRIHWEVNTWRMHYSEEMTQISMLVLTGAFFH 600  
QY 807 KYVAYLYSVYVPOAIRDMVDEYINCEDIANFLVSHTRRPPIKVTSRMTFFRPGC--PQ 864  
DB 601 KYVSHYTHAMPQDIDKWDVDEHNCEDIANFLVANTTNPPRIKYVTRPKKFKCEPCANTE 660  
QY 865 ALSHDSHFHERHKCINFEVKKYGYMPLTYTOPRVDSVLEFKTRLPHDKTKCFKFI 919

DB 661 MLSADLNMRERSACIDRFESKTYGRMPLRTVEFRADPVLFRDNPP-DKIKRYNDI 714

## RESULT 11

ID 098SV5 PRELIMINARY; PRT; 738 AA.

DT 01-JUN-2001 (TREMblrel, 17, Created)  
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)  
DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)

## DE EXOSTOSIN.

OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
NCBI\_Taxid=8355;

## RP SEQUENCE FROM N.A.

RA Hill A., Brown N., Hill M., Wells D.E.;  
RT "Identification of a new ext family member in Xenopus laevis."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF319538; AAK15278.1; -  
DR InterPro: IPR004263; Exostosin.  
DR Pfam: PF03016; Exostosin; 1.  
SQ SEQUENCE 738 AA; 84955 MW; D6F68E65720D085 CRC64;

## Query Match

Best Local Similarity 17.1%; Score 832; DB 13; Length 738;  
Matches 244; Conservative 123; Mismatches 236; Indels 232; Gaps 35;

QY 150 QPKLSLP-IRLLPEKDDAGLPP-----PKATRG-----CRLHNCDFYSCPLTSG 193  
DB 47 QPWLILPDTSLPPTQEGGDPGQSVPRQKROSNSSITYGKRRMSCDLSIC-RRNG 105  
QY 194 FPVYVYVSDQFVFGSYLDPLVKAQAFQATARAANYVTENADIACYVILVGEVMOEPVYLR 253  
DB 106 FKVIYIPQOK---GDKLSDSYQHILAIETISRFY-TSDPSQACVFLSLDTLDRD-QLSP 160  
QY 254 ---AELKOLYSLPHMTDGNHNTINL---SRKSDPQNLLYNSVGRANVAOSTFTYVQ 307  
DB 161 HYVHNLSKVQONLHLW-NNGRNHLIFNLVSGTWPDTYEDVFDI--GQAMLAKASISTEN 217  
QY 308 YRPGFDLVSPVYHAMSEP-----NMEIIPQVYVRYKYLTFPGCEKI-----EST 353  
DB 218 FRPWFQD-VSLPFLSKHPRFGDGKGLFRNNNIPM---RKVLYVKKGRYLTIGISDT 271  
QY 354 RSSIQEAKSFEEEMEGDPADYDRIITATLKAVODSKLDQVLYEFTCKNPKPSLPTFW- 412  
DB 272 RNALYHVIMGED-----VLLTTCCKHGKK-----WQ 297  
QY 413 ---ALCG-----EREDRLLEKLSTFALITTPGDRLVISGCAATREALEVGAIVP 462  
DB 298 KHKDARCDKDAEYKEDYEMLEHNAATFCV--PRGRRL-----GSRFLLEALQACV 350  
QY 463 VLGEVOLPYODMLQWNEALVVPKRYTEVHLLRSISDSDDLAMRGQRFMETYFST 522  
DB 351 MLSNGWELPSEVYIDWQAIVIGDERLLQIPSTIRSIHQDKLIALRQVQFIMEAVFSS 410  
QY 523 ADSIEFNTVLMITRQIOIPAPRIREAAAEIPIRSGKAAGTDPNNADNGDLGPEVTEP 562  
DB 411 VERIVTLTTELIQDRI-----FKHIS----- 431  
QY 583 PYASPRILRMFTLVTDVFRSNCAPGFHLPHTPPDPVLPSEAKFLG-----SGTGF 636  
DB 432 -----RNSLM-----WKKHFG--GLF-----VLPOYSTYLGGFPYVYNTLGA 466  
QY 637 RPIGGAGSGGKEFOA-----LGGNVPREQFTVVMLTREEREVLNMSLERLNGLPYIN 690  
DB 467 KP-----PSKFTKCDPCPCNPPGYSVPTNPPAAMCCTKSQ-----CWYIN 505  
QY 691 KVVVVVWNSPK-LPSEDDLMPDIGVPIVYVTRTEKNSLNRLPWNIEFEALISIDDADHL 749

```

Db 506 --TVLMWCDPFLPAKH-RWPATTVPVIVIEGSKVSSRFLPDNDGTDAVLSDDETVL 562
QY 750 RHDEIMGFVREARDRIYGFPGRYHAMDIPIHOSMLYNSYCELSMTLTAAGAFHKY 809
Db 563 STEVEDPAFTVWOSFPRIVGYPARSHFMDSTERNQYTSKWTNDMSVLTGAITHKY 622
QY 810 AYLXSYVMPQALIDMVDYINCEDIANMLVSHITRKPRIKVT-----S 853
Db 623 HYLYTNVLPALSKNMVDQLANCEDILMNFVLSAVTKLPRIKYOKQYKETMNGOFFRAS 682
QY 854 RMFRRCGCCQALSHDSSHFRHKCNFVVKYGYRPLLYTQERDVSLEFKRL 908
Db 683 RMA-----DPDHAQROTCLNTFAWTGLGYMPLLSQMRDLPVLEKQV 725

RESULT 12
076796 PRELIMINARY: PRT: 760 AA.
076796;
01-NOV-1998 (Tremblrel. 08, Created)
01-NOV-1998 (Tremblrel. 08, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE TUMOR SUPPRESSOR HOMOLOG.
GN TTV OR TOUT-VEU OR CG10117.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98328070; PubMed=9665133;
RA Bellaique Y., The I., Perrimon N.;
RT "Tout-veu is a Drosophila homologue of the putative tumour suppressor
RT EXT-1 and is needed for hh diffusion.";
RL Nature 394:85-88(1998).
DR EMBL: AF083889; AAC32397.1; -.
DR FlyBase: FBgn0020245; tlv.
DR InterPro: IPR004263; ExoScotin.
DR Pfam: PF03016; ExoScotin; 1.
SQ SEQUENCE 760 AA: 87037 MW: EE48C040B52B54DB CRC64;

Query Match 16.3%; Score 796.5; DB 5; Length 760;
Best Local Similarity 27.0%; Pred. No. 3,6e-55;
Matches 228; Conservative 129; Mismatches 284; Indels 205; Gaps 26;

122 ENMK-----QDLIDKNIISTEHSYKELMAQNOPK-LSLPIRLLPKDDAGLPKATRG 176
QY 40 ESAKDGCVQPHEDLPFLGAHMQEQLLOLSNOSKSLDSSKHLVTRKPD----- 88
Db 177 CRHNCDFYRCPDLSGFPYVVDQFVGSLDPLVKAQFOATARANY----- 226
QY 89 CRMETGCDPFRCL--YDRFLVYIYPP-----EPLNSGAAPRPSANTOKLITAIQES 137
QY 227 -VYTEMNADICLVILVIGEMQEPVVLNP--AEDEKOLYSLPHMRTDGHNVIIINLSKRS- 282
Db 138 RYTSDEPTACLFVLGIDTLDRDLSLSEDYRVNPSRLARLPYV--NNGRNHIIINLXSGTW 196
QY 283 -DTONLLIYNSTGRANVASTFTTVQVRPFDLVNSPLVH-----ANSEPFME 330
Db 197 PDVAENSLGFDAGEALILAKSMGVLLRHCQFVSI--PLFHKQPLRAGATGVQSNMF-- 253
QY 331 IPROVPKRYKRYLFTFOGEKI-----ESLRSSLOEARSFEEMGDDPADYDDRIATL-- 383
Db 254 -----PANKRYLLAFKCKRIVHIGSETNSLPHLNGR-----DNLVYTTCR 296
QY 384 -----KAVODSKLDQVLEVFCTKNOPKSLPTEMALGGERDRLELKLSTFALIITPGD 438
Db 297 HGKSMRELQDNCRDEDEHREX-----DAYDETLLQNSTFCLV--PRG 336
QY 439 PRVYSSGCATRLPEALEVGAPEVVLGEVOVLYQDMQLQWNEALVVPKPRVTEVHFLR 458

```

```

Db 337 RRL-----GSFRLEALQAGCIPVLLSNAMAVLPEESKIDMKQAIAWDELLQLQVPIVR 391
QY 499 SLSDSDLLAMRRQGRFLMETYFSTADSIENFTVLAAMI RTRIQIAPAIIRERAAEIHRS 558
Db 392 SIPEAKIFALROQTQVLMERYFGSIEKIVFTTEILIRER--LPDYVRSSLV----- 441
QY 559 KAAGTDPNMANDGLDGLPEVEEPYASPRYLRLNTLYTDFEYSNMCAPGFHLPP-HT 617
Db 442 -----NNSPGALLLTPTRA 456
QY 618 PEDPVLPEAKFLGSGTGFPFGCGAGSGCKEFOALGCVNPEQPTVMVLTIREBEVLM 677
Db 457 DSSRYMP-----FLNLSMGAEPRHNYTAVIYVQGAALGPA-----ALY 496
QY 678 NSLERLNGLPYLNKVVVWNSPK-LPSEDLMPDIDGPIVNV-----RT 720
Db 497 KLVRTITKSGFVERILVLAADRPDLKRRMPTSHIPLHVISLGSTRSGAGPTSOIT 556
QY 721 E-KNSLNRLPWNIEIETEAISIDDAHLRHDEINFGFVRREARDRIYGFPGRYHAND 779
Db 557 EGRPSISQRFPLPYDEIQTDAVLSDDEALINTDELDPATVYVMDPFRIVGYPARAHFMD 616
QY 780 IPIHOSMLYNSYCELSMTLTAAGAFHKYIAYLYSVMPQALIDMVDYINCEDIANML 839
Db 617 DSKNAGYSKMTNYSIVLTGAFFHYRNYLYTWMLSLLKTKVQOSSNCEDIIMNLL 676
QY 840 VSHITRKPRPKVSRMTFRCPQALSHDSSHFRHKCNFVVKYGYRPLLYTQERDVSLEFKRL 899
Db 677 VSHYTRKPRPKVYQKRGLOGSGDSGFAMNDPHFIQROSCLTNFAAVGYMPLIRSLRN 736
QY 900 DSVLEFR 905
Db 737 DPMLYR 742

RESULT 13
09V730 PRELIMINARY: PRT: 760 AA.
ID 09V730;
AC 09V730;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TTV PROTEIN.
GN TTV OR CG10117.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richardson S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Moharir C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Shue B.C., Siden-Klamos I., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Slater E., Spradling A.C., Stapleton M., Skupski M.P., Smith T.,  
 RA Styrtas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003814; AAF58236.1; -  
 DR FlyBase: FBgn0020245; tlv.  
 DR InterPro: IPR004263; Exostosin.  
 DR Pfam: PF03016; Exostosin: 1.  
 SQ SEQUENCE 760 AA; 87308 MW; F6369FDBA206DEA9 CRC64;

## Query Match

16.3%; Score 795.5; DB 5; Length 760;

Best Local Similarity 26.8%; Pred. No. 4,4e-55;

Matches 227; Conservative 132; Mismatches 282; Indels 205; Gaps 26;

QY 122 ENAK-----QDLQKNVISQTEHYKELMAONPK-LSLPRLPEKDDAGLPPPKATRG 176  
 DB 40 ESARDGVQPHEDLPSEFLGANDMOELQLONSKSLDSSKHLTKRPD----- 88  
 QY 177 CRLNHCEDYSRCPLTSGEPYVYDSDQFVFGSYLDPLVKAQFQATARAVY----- 226  
 DB 89 CRMETCTDFTRC--YDFELVYIIRP-----EPLNSLGAAPPTSNVCKITLTAQES 137  
 QY 227 -YVTENADIACLVYLVGEMQEPVLRP--AELEKOLYSLPHMRTDGHNVIMLSKRS- 282  
 DB 138 RYYSIDPTACLFYLGIDTLDRDLSLSDYVYRVNPSRLALRPYW--NQRNHIIFNLVSGTW 196  
 QY 283 -DTQNLINYNSTGRAMAQSFTYVQYRPGFDLVVSPLVH-----AMSEPNME 330  
 DB 197 PDVAENSLGFDAGAILAKASMGVLQLRHGDSVI-PLFHKQFPLRAGATGYOSNFF-- 253  
 QY 331 IPPQVAKRYLYTFQGEKI-----ESLRSLQEARSFEEEMEGDPADYDDRIATL-- 383  
 DB 254 -----PANKKTYLAFKRRKRYVHGSETKNSLFLHNGR-----DMVLVTTCR 296  
 QY 384 -----KAVQDSKLDQVLVEFTCKNPKRPSLPTEMALCGEREDLELLKSTFALIITPGD 438  
 DB 297 HGKSMRELQNRCEDEDEREY-----DRDYETLLQNSFCLV--PRG 336  
 QY 439 PRVYISGCGATRLFEALGVAVPVYLVGEQVQVLPYQDMLQNNALVYPKRYVEVHLLR 498  
 DB 337 RRL-----GSFRLEALQACIPLVLSNANVLPRESKIDMKQAIAWADERLLQVPIVR 391  
 QY 499 LSLSDSLAMRROGRFLMETEYFSTADSIPTVLMIRTRIQIPAPIREBAAEIPIHRS 558  
 DB 392 SIPIERFALRQGTQVLEMEYFSGIEKIYFTTEILIRER--LPDYPRSSLY----- 441  
 QY 559 KAAGTDPNMAKDGLGVETEPYASPRYLKRNFTLLVTDFYKSNWCAQGPFLRP-HT 617  
 DB 442 -----WNSSPGALLTLTPFA 456  
 QY 618 PFDVLPSEAKFLGSGTGRFPIGGAGSGKKEQALGAVNPRQGVVWMLTYEREVLM 677  
 DB 457 DSSRYMP-----FLNSKGAEPRHNTAVIYVQAGALGPNA-----ALY 496  
 QY 678 NSLERLNGLPYLNKVVVWNSPK-LPSEDLILPDIQVPIWV-----RT 720  
 DB 497 KIVRTITKSGFERIIVLWMAADRPLPLKRWPRPSHIFPLHVISLGSGTRSGAGPSTQTT 566

QY 721 E-KNSLNNRFLPMWEITEALISIDDAHLRHEIMGFRVWRARDRIVGPGRYHAMD 779  
 DB 557 EGRSISGRFLPYDEIQDVALSDEDAILNTDELDFAYVWRQPEPIVGYPARAHWD 616  
 QY 780 IPHOSWLYNSYSCSELSWLTGAAPFHKYAYLYSYVPOAIRDMVDEYINCEDIAMN 839  
 DB 617 DSKNAMGYSKWTYYSIVLTGAAFYHRYNYVLTNNMLSTLLKTVQOSSNCEDIKMLN 676  
 QY 840 VSHTRKPKIVTSRMWFRCPGQALSHDSHERHKCINFPVKYGVWPLLYTQFRV 899  
 DB 677 VSHTRKPKIVTKRKQKIDREGRSPWDPDHFIOQSCUNTPAAVFGIWPILIRNLM 736  
 QY 900 DSVLEK 905  
 DB 737 DPMLYR 742

## RESULT 14

Q9JRV7 PRELIMINARY; PRT; 669 AA.  
 ID 09JRV7;  
 AC 09JRV7;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE MULTIPLE EXOSTOSES-LIKE 1 PROTEIN.  
 GN EXTL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stickens D.J., Brown D., Evans G.A.;  
 RT "Ext genes are differentially expressed in bone and cartilage during  
 RT mouse embryogenesis,"  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF224461; AAF61913.1; -  
 DR MGI:1888742; Extl1.  
 SQ SEQUENCE 669 AA; 74008 MW; CD547AEF5B4EAB1 CRC64;

## Query Match

13.1%; Score 639; DB 11; Length 669;

Best Local Similarity 26.1%; Pred. No. 1,4e-42;

Matches 203; Conservative 117; Mismatches 221; Indels 238; Gaps 29;

QY 171 PKATRG--CRLHNFQDSRCPLISGFYVYDSDQFVFGSYLDPLVKAQFQATARAVYV 228  
 DB 69 PRYSRGSCTWGAQCFDYSKCR-----GVLKIFVHSPAGPTSEAQRRI 111  
 QY 229 TEN-----ADIACLVYLVGEMQ-----PVYLRPAELEKOLYSLPHMRTDGHNH 273  
 DB 112 LDSLEGRYSALSPAD-ACLLFLPSQDRACGCLP-----PNW--NGGRNH 156  
 QY 274 VIINLSKRSQTONILYNVSGRAMVASTFTYVQYRPGFDLVVSPLVHMANSEPNMEIIP 333  
 DB 157 LVLSLPAPCTR-----LQAMVVAEASPSDIFRFGFIAL-----P 193  
 QY 334 QVPKRYLYTFQGEKIESLRSSLOEARSFEEEMEGDPPA---DYDDRIATLKAQVDSK 390  
 DB 194 YLP-----EA-----HPLRGAPRKIQHSHQRPATLTLAAAEK 227  
 QY 391 -----LDQVLVEFTCKNPKRPSLPTEMALCGEREDLELLKSTFALIITPGDRL 441  
 DB 228 GRWRITSTHASACIMRCHQDQGP-----QDTYGETLIPNATFCLL--PG----- 271  
 QY 442 VISSGCTRLF-EALEGANPVYLVGEQVQVLPYQDMLQNNALVYPKRYVEVHLLRSL 500  
 DB 272 ---HRSATSCFIALQACIPLVLSRWELPFESEVIDMTAAIADDERLPLOYLAALREM 328  
 QY 501 SDSDDLAMRROGRFLMETEYFSTADSIPTVLMIRTRIQIPAPIREBAAEIPIHRS 560  
 DB 329 LPSRVIALRQGTQVLEMEYFSGIEKIYHTTLEVIQDRIW-----G 368

QY 561 ACTDPNNMADGDLDCGVEPEPPYASPRYLKNTLVTFDFRSMNCAPGP-HLEPHTPF 619  
 Db 369 ASHPSFSLMWN-----SPGAL-LALPTFSTSLQDF-----PFYH----- 402  
 QY 620 DPLVPSAKFLGSGTGRPIGGAGSGCKEFOALGNNVREOFTVYMLYEREEVLMS 679  
 Db 403 -----QLGSG-----PGSSFSAYIWWGASGESLKL 428  
 QY 680 LERLNGLPYLNKYVWNSKPLSPEDLMPDIPYIMVYTERKNSLNRPNNIETEA 739  
 Db 429 IOEYAGSRHCAQLILILNSKPPPP--RMEPIVPLVIGHR-KVSNRFFYSNISTNV 485  
 QY 740 ILSDDAHLRHDEIMGFVWREARDRIVGFGRYHAMDIPHOSWLYNSNYSCELSMVL 799  
 Db 486 ILSDAOSTLSTSEVDFAFVWQSFPERMVGFLSGSHFDEAOGMGVRTGMTNDFSMVL 545  
 QY 800 TGAFFFKYAYLYSVYMPQAIQMDVEXINCDIAMNPLVSHITRKPIKYTSRWTFRC 859  
 Db 546 TTAFFHYHYHTLFTSHLPKALRTIADPTCVDLNPLVATVTKLPKIKVP--YGRH 603  
 QY 860 PGC-----POLSHDSDHFERHKCINFEVYGYMPLLYTOFRVDSVLFK 905  
 Db 604 PEAVPMDSGPRPRVPERQPLDQ-----CINRLAAGFGHPLVSSQVRDPLVLFK 653

RESULT 15

ID Q9D7G5 PRELIMINARY; PRT; 131 AA.  
 AC Q9D7G5;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY.  
 DE CLONE:2310009010, FULL INSERT SEQUENCE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustlincich S., Hill D., Hornann W., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK009257; BAB26174.1;  
 SO SEQUENCE 131 AA; 14976 MW; E8ED3DB7BB40A7 CRC64;

Query Match 12.0%; Score 586; DB 11; Length 131;  
 Best Local Similarity 96.6%; Pred. NO. 2e-39;  
 Matches 112; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MTGYTMRNGAGCGGTCMLRNSNRIRLTWLSFTLFIIVFFPLAHYYLTTLDEADEA 60  
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

Db 1 MTGYTMRNGAGCGGTCMLRNSNRIRLTWLSFTLFIIVFFPLAHYYLTTLDEADEA 60  
 QY 61 GKRIFGPRVGNELCEYKHYLDLCRIRSVSEELLQLEAKROELNSEIAKLNTKTEA 116  
 ||||||| :||||||||| ||||||||||| ||||||||||| |||||||||||  
 Db 61 GKRIFGPRAGSELCEYKHYLDLCRIRSVSEELLQLEAKROELNSEIAKLNTKTEA 116

Search completed: September 4, 2002, 15:02:40  
 Job time: 299 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 17:58:59 ; Search time 4303.49 Seconds  
(without alignments)  
19357.128 Million cell updates/sec

Title: US-09-809-920-3  
Perfect score: 6172  
Sequence: 1 GGGCGGTCCCTGACCTGGA.....ATTAATGAGACTTTAACTC 6172

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
arched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	904.8	14.7	964	9	AL520206
2	893.4	14.5	928	9	AL520207
c 3	870.6	14.1	919	9	AL522291
4	844.6	13.7	871	9	AL522292
5	810.4	13.1	874	9	AU131026
6	764.8	12.4	803	9	AU125122
7	736	11.9	769	10	B1554177
8	734.4	11.9	770	10	AU140433
9	725.8	11.8	743	9	AU140433
10	715	11.6	763	10	B1196587
11	706.4	11.4	734	9	AU132259
12	706	11.4	768	9	AU133145
13	672.4	10.9	1000	10	BG115714
14	666.2	10.8	755	9	AU133571
15	663	10.7	941	10	BF982138
16	662.4	10.7	740	10	BM050902
17	658.6	10.7	968	10	BG424375

c 18	650	10.5	672	10	BE343959	BE343959	602015653
19	642.4	10.4	644	12	AQ334883	RPCT-11-3	
20	637.2	10.3	885	10	BG423077		
21	636.4	10.3	775	9	AL528972	AL528972	AL528972
22	634.2	10.3	987	10	BG023953		
23	633.4	10.3	635	10	BC910960		
24	633	10.3	644	9	AW603974		
25	628.8	10.2	644	9	AW603972		
26	620.4	10.1	736	10	BG261277		
27	617	10.0	853	10	BF311835		
c 28	598.8	9.7	627	10	BE839880		
29	596	9.7	804	10	BI911162		
30	592.4	9.6	623	9	AW954438		
31	589.2	9.5	694	10	BM047368		
32	588	9.5	588	10	BE879757		
c 33	581.4	9.4	658	12	AG165871		
34	570	9.2	957	10	BE897823		
35	560.8	9.1	742	10	BF316353		
36	557	9.0	557	9	AL046695		
37	556	9.0	616	10	BE264316		
38	550.6	8.9	579	10	BF311163		
39	546	8.8	569	10	BE383505		
40	543	8.8	595	9	BE142805		
41	538.6	8.7	593	10	BE265859		
42	534	8.7	546	10	BE278303		
43	531.8	8.6	704	10	BG924082		
c 44	529.8	8.6	559	9	AU148856		
45	529	8.6	529	10	BF725639	bx1v1e11.y	

## ALIGNMENTS

RESULT 1  
AL520206/c  
LOCUS AL520206 964 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL520206 LTI\_NFL004\_NBC2 Homo sapiens CDNA clone CSDB005YN16 3  
prime, mRNA sequence.  
ACCESSION AL520206  
VERSION AL520206.1 GI:12783699  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 964)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

source  
1. 964  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDB005YN16"  
/clone.lib="LTI\_NFL004\_NBC2"  
/sex="male"  
/tissue="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email: fliang@lifetech.com URL : http://fulllength.invitrogen.com" 5 others  
BASE COUNT 227 a 239 c 272 g 161 t



OY	5252	TATTCCTTTTACAGTAGCCTTCTCCCTCGGGGAGCCAAAGAAGTGCTGTGTGGCCT	5311
Dd	361	TATTCCTTTTACAGTAGCCTTCTCCCTCGGGGAGCCAAAGAAGTGCTGTGTGGCCT	420
OY	5312	AATAATTGGCTCTATTTTCATCATCGTGTTCCTTTTAATGTAGGAACATCAATACTGCTTC	5371
Dd	421	AATAATTGGCTCTATTTTCATCATCGTGTTCCTTTTAATGTAGGAACATCA - AATCTGACTTC	479
OY	5372	ACTGGAGCTCGGTGAGACC GG GGCCTGTGTGTGTGGGACCCCTTTAAGCGGACTCA GT	5431
Dd	480	ACTGGAGACTCGGTGAGACC GG GGCCTGTGTGTGTGGGACCCCTTTAAGCGGACTCA GT	539
OY	5432	GAGCTGGGGGCCCTGTGTGTGTGGGAGCCAGGGCCCTCTCTTACTGGAGCACAGGTTGTG C	5491
Dd	540	GAGCTGGGGGCCCTGTGTGTGTGGGAGCCAGGGCCCTCTCTTACTGGAGCACAGGTTGTG C	599
OY	5492	GGGCCCCGCAATGTCACTGGTGGATTAAGAAAGGCGTGA GTGGTGTGACACCAAACATGC	5551
Dd	600	GGGCCCCGCAATGTCACTGGTGGATTAAGAAAGGCGTGA GTGGTGTGACACCAAACATGC	659
OY	5552	CGCAGGAGAGCGCTGTGTGTGCGCGGTCTTCCAACAAGACAGCCCTCTTGACCTTGAAG	5611
Dd	660	CGCAGGAGAGCGCTGTGTGTGCGCGGTCTTCCAACAAGACAGCCCTCTTGACCTTGAAG	719
OY	5612	GAACACTGGCTTGAAGAGACTGCAGACAGGCTGTGAAGGGGACGGCCCTCTCAAGGAGAGG	5671
Dd	720	GAACACTGGCTTGAAGAGACTGCAGACAGGCTGTGAAGGGGACGGCCCTCTCAAGGAGAGG	779
OY	5672	CAGCAAGGTGGCCACAGTGTCACTAGTCAAGTGTCTTCCACACAGGAGGAAAGCGCGGACC	5731
Dd	780	CAGCAAGGTGGCCACAGTGTCACTAGTCAAGTGTCTTCCACACAGGAGGAAARCGCGGACC	839
OY	5732	TGTGACTGCTTTGAAGTGGGAAAAGCGGCGCCACAACCCCGGCTCTCTTGGCTGTGTGT	5791
Dd	840	TGTGACTGCTTTGAAGTGGGAAAARCGGCGCCACAACCCCGGCTCTCTTGGCTGTGTGT	899
OY	5792	GGGCGCGCCCTGTGGACCTTGTCTGGGCTCG	5822
Dd	900	GGGCGCGCCC--TGGCACTTGTCTGTGGCTCG	928
RESULT	3		
LOCUS	AL522291/C		
DEFINITION	AL522291.LTI_NFL004.NBC2 Homo sapiens cDNA clone CSDB008YN08 3		
ACCESSION	AL522291		
VERSION	AL522291.1 GI:12785784		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	l1 (bases 1 to 919)		
JOURNAL	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.		
COMMENT	Full-length cDNA libraries and normalization unpublished (2001)		
FEATURES	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers 1..919 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CSDB008YN08" /clone_1lb="LTI_NFL004.NBC2" /sex="male" /tissue.type="neuroblastoma cells" /lab.host="DH10B"		
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Dll9(dT) primer. Five prime end		

Query Match	Best Local Similarity	Score	DB 9	Length	919
Matches	903	Conservative	8	Mismatches	7
				Indels	3
				Gaps	3
5214	AGACCTTCTAGTGGCCCTCTACCTGACTCTGACTTGTATTCTCTTTTAGCAGTACGCTT	5273			
919	AGAACCTTCTAGTGGCCCTCTACCTGACTCTGACTTGTATTCTCTTTTAGCAGTACGCTT	860			
5274	CTTCCCTGGGGAGGC-AAAGAGTGGGTGTGGGGCTATTTTGGGTGGCTATTTCAT	5332			
859	CTTCCCTGGGGAGGCMAAGAGTGGGTGTGGGGCTATTTTGGGTGGCTATTTCAT	800			
5333	CTGGTTCTCTTTTAAATGTAGAACTACATACTGACTTCAATGGGACTCGTAGCCGGG	5392			
799	CTGGTTCTCTTTTAAATGTAGAACTACATACTGACTTCAATGGGACTCGTAGCCGGG	740			
5393	GGCGTCTGTGTGGTGGGAGCCCTTTAGGGGAGTCAGTACTGGGGCGCTGTGTGG	5452			
739	GGCGTCTGTGTGGT-GGACCCCTTTAGGGGAGTCAGTACTGGGGCGCTGTGTGG	681			
5453	TGGAGCCAGAGGCGCTCCCTTTAGTGAAGCCAGTGTGTGGGCCCGCGAATTCACATGGTG	5512			
680	TGGAGCCAGAGGCGCTCCCTTTAGTGAAGCCAGTGTGTGGGCCCGCGAATTCACATGGTG	621			
5513	GATCTAAGAGGGCTGAGTGTGTGTGACACCAAAACATGCCCCAGAGGAGGCTGTGTGCC	5572			
620	GATCTAAGAGGGCTGAGTGTGTGTGACACCAAAACATGCCCCAGAGGAGGCTGTGTGCC	561			
5573	GGTGTCTTCAAAAGAGAGCCCTCTTACCCCTGAAAGAACACTGGCTTTAAAGACTG	5632			
560	GGTGTCTTCAAAAGAGAGCCCTCTTACCCCTGAAAGAACACTGGCTTTAAAGACTG	501			
5633	CAGACAGGCTGTAGAGGGGACGCCCTCTCAGCGAGAGGACAGCAAGTGGCCACAGTGTG	5692			
500	CAGACAGGCTGTAGAGGGGACGCCCTCTCAGCGAGAGGACAGCAAGTGGCCACAGTGTG	441			
5693	ACTGGTCAGGCTCTCTCAACAGGGGAAAGCGCGCACTGTGACTCGCTTGAAGTGGGA	5752			
440	ACTGGTCAGGCTCTCTCTCAACAGGGGAAAGCGCGCACTGTGACTCGCTTGAAGTGGGA	381			
5753	AAGCGGCGCAGACAGACCCCGGGGTCTCTTGGCTGTGTGTGGGCGCGCCCTGGCCACTTGG	5812			
380	AAGCGGCGCAGACAGACCCCGGGGTCTCTTGGCTGTGTGTGGGCGCGCCCTGGCCACTTGG	321			
5813	TCTTGGCTGTGAGGTGACAGAGCGGCTGTCTCTGGTGGCCGGGCTTGTGTCTCGGT	5872			
320	TCTTGGCTGTGAGGTGACAGAGCGGCTGTCTCTGGTGGCCGGGCTTGTGTCTCGGT	261			
5873	TTCGGGTCTTACCATTAACACCGTCCAGGGGCTGTGACAGGGCAGTGTAGGGCTGGATC	5932			
260	TTCGGGTCTTACCATTAACACCGTCCAGGGGCTGTGACAGGGCAGTGTAGGGCTGGATC	201			
5933	CTTGGGCACTGCTCTCCGTGTGACTGTGCTCAGGCGCCAGGGCTACACAGTGGGGTCC	5992			
200	CTTGGGCACTGCTCTCCGTGTGACTGTGCTCAGGCGCCAGGGCTACACAGTGGGGTCC	141			
5993	TCTTGGGAAAGATGGGATTTCTTGGAGGCTGGCGGACAGAGTGGGGAGTCTCCAGTT	6052			
140	TCTTGGGAAAGATGGGATTTCTTGGAGGCTGGCGGACAGAGTGGGGAGTCTCCAGTT	82			
6053	TCTTGGGAAAGATTTGATATCATATGCCACGTCCATCCACCCACCCCTTTTCTGTACAGA	6112			

```

Db      81  TGTGGGGGAGGCTTTGATATCATGCGACGCCACCCACCCCAATACGTMACGA 22
QY      6113 GCACAAATGCTGTACATTTGA 6133
Db      21  GCACAAATGCACTTACATTTGA 1

RESULT  4
AL522292 871 bp mRNA linear EST 13-FEB-2001
LOCUS    AL522292 LTI_NFL004_NBC2 Homo sapiens cDNA clone CSDB008YN08 5
DEFINITION Prime, mRNA sequence.
ACCESSION AL522292
VERSION   AL522292.1 GI:12785785
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 871)
AUTHORS   Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 Evry cedex - France
           Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source
1. 871
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="CSDB008YN08"
   /clone_1lb="LTI_NFL004_NBC2"
   /sex="male"
   /tissue_type="neuroblastoma cells"
   /lab_host="DH10B"
   /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
   was primed with a NotI-oligo(dT) primer. Five prime end
   enriched, double-stranded cDNA was digested with Not I and
   cloned into the Not I and Eco RV sites of the pCMVSPORT 6
   vector. Library was normalized. Library was constructed
   by Life Technologies. Contact: Feng Liang Life
   Technologies, a division of Invitrogen 9800 Medical Center
   Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
   8371 Email: fliang@lifetech.com URL :
   http://fulllength.invitrogen.com"

BASE COUNT 204 a 200 c 246 g 220 t 1 others
ORIGIN

```

```

Query Match 13.7%; Score 844.6; DB 9; Length 871;
Best local Similarity 99.9%; Pred. No. 9,7e-195;
Matches 844; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGGGGGCTCTGAGCTGGAAGCGGAGAGCAAGCCCTGAGGTTCACTTTTCAAGAGT 60
Db      27  GGGGGGCTCTGAGCTGGAAGCGGAGAGCAAGCCCTGAGGTTCACTTTTCAAGAGT 86
QY      61  CGGTGCTGAGGTGTATGCTACACAAGTGAAGGAAGAGGCTCTGAAACATGCG 120
Db      87  CGGTGCTGAGGTGTATGCTACACAAGTGAAGGAAGAGGCTCTGAAACATGCG 146
QY      121 CTGATTGTTGGCAAGGATCATAGAAGCTGCAATTTATTTCTGTTCACTTACT 180
Db      147 CTGATTGTTGGCAAGGATCATAGAAGCTGCAATTTATTTCTGTTCACTTACT 206
QY      181 GTATTAAGTGTGAATAGACATATTTGTTGTACAAACCAAGAACAGAGAGC 240
Db      207 GTTAACTGTGAATAGACATATTTGTTGTACAAACCAAGAACAGAGAGC 266
QY      241 TATGCAATTTGAAAAGTCTGTGATTCAGAGGCTTTTCTGGGTTTCATCATCAG 300
Db      267 TATGCAATTTGAAAAGTCTGTGATTCAGAGGCTTTTCTGGGTTTCATCATCAG 326
QY      301 TACCTCCCTCCCTTCATCTCAGCAAGAAATGCGACCTTTTATGCTTGTATTAAGATTAA 360
Db      327 TACCTCCCTCCCTTCATCTCAGCAAGAAATGCGACCTTTTATGCTTGTATTAAGATTAA 386
QY      361 GGACATGTTCTTTGGTCAACAGCAGAACTTAAATCTGCTGGAATAGGCTCAGAGACA 420
Db      387 GGACATGTTCTTTGGTCAACAGCAGAACTTAAATCTGCTGGAATAGGCTCAGAGACA 446
QY      421 TTTGAGCTGCAAGCTGAGGAAAATGAATGTTCAATTTATTTTGGTCTGCTGGGAGC 480
Db      447 TTTGAGCTGCAAGCTGAGGAAAATGAATGTTCAATTTATTTTGGTCTGCTGGGAGC 506
QY      481 ACACCTAAGCTTTCTGGAACAGCTGCAAGAGAGATGCTTTTGTGAATAGCAACC 540
Db      507 ACACCTAAGCTTTCTGGAACAGCTGCAAGAGATGCTTTTGTGAATAGCAACC 566
QY      541 ATGCTATGCGAGTGAACCCGACGCTGATCTGGGGGAGAGCTGCAGAGACTCATGACAG 600
Db      567 ATGCTATGCGAGTGAACCCGACGCTGATCTGGGGGAGAGCTGCAGAGACTCATGACAG 626
QY      601 GCTATACCATGCTCGGAAATGGGGGCGCGGGAACGAGATGCACTGCTGCGCT 660
Db      627 GCTATACCATGCTCGGAAATGGGGGCGCGGGAACGAGATGCACTGCTGCGCT 686
QY      661 GGTCACACCGCATCGGCTCGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db      687 GGTCACACCGCATCGGCTCGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
QY      721 TCCGCGTATCGCCGCTATTTACCTCACCACCTGATGAGGCTGATGAGGAGCAGCAGC 780
Db      747 TCCGCGTATCGCCGCTATTTACCTCACCACCTGATGAGGCTGATGAGGAGCAGCAGC 806
QY      781 GGATTTTGGTCCCGGGGTGGGAGACAGCTGTGCGAGGTGAAGCAGCTGCTGATCTGT 840
Db      807 GGATTTTGGTCCCGGGGTGGGAGACAGCTGTGCGAGGTGAAGCAGCTGCTGATCTGT 866
QY      841 GCCGC 845
Db      867 GCCGC 871

RESULT  5
LOCUS    AU131026 874 bp mRNA linear EST 24-OCT-2000
DEFINITION AU131026 NT2RP3 Homo sapiens cDNA clone NT2RP3001873 5', mRNA
sequence.
ACCESSION AU131026
VERSION   AU131026.1 GI:10991380
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 874)
AUTHORS   Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
           Yamamoto, J., Makatsuta, A., Nakamura, Y., Nagai, T., Sugano, S. and
           Isigai, T.
TITLE     HRI human cDNA project
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3951
           Fax: 81-438-52-3952
           Email: genomics@hri.co.jp
           HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
           Research Institute; cDNA library construction: Department of
           Virology, Institute of Medical Science, University of Tokyo, and
           Helix Research Institute.
FEATURES
Location/Qualifiers

```

```

source 1.874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3001873"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/line="NT2"
/notes="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 176 a 221 c 219 g 254 t 4 others
ORIGIN

```

```

Query Match 13.1%; Score 810.4; DB 9; Length 874;
Best Local Similarity 97.3%; Pred. No. 2.1e-186;
Matches 854; Conservative 0; Mismatches 20; Indels 4; Gaps 3;

```

```

3691 TATTTAAAAAGAAAGTTTCAATTTGCCATTCAGAGCTTATTTATATATAGTGTG 3750
|||||
1 TATTTAAAAAGAAAGTTTCAATTTGCCATTCAGAGCTTATTTATATATAGTGTG 60

```

```

3751 TATATAATACATGACACACTTGATATATATTTTGGCTGGGAGTGTAGTT 3810
|||||
61 TATATAATACATGACACACTTGATATATATTTTGGCTGGGAGTGTAGTT 120

```

```

3811 TTGCTTTCTAAGGAGGAGCGCGAGGCTCTTGTCTGTATTCGTGGGAGATGG 3870
|||||
121 TTGCTTTCTAAGGAGGAGCGCGAGGCTCTTGTCTGTATTCGTGGGAGATGG 180

```

```

3871 TTCTGGCTGTGTCACTGGCTTATCTTAAATCATCTCCATCTCCCGAGCCAT 3930
|||||
181 TTCTGGCTGTGTCACTGGCTTATCTTAAATCATCTCCATCTCCCGAGCCAT 240

```

```

3931 CTGTGTGCAACAACGAAAGGATGAATTCGCTTGGCGGCTGGACAGGCTTC 3990
|||||
241 CTGTGTGCAACAACGAAAGGATGAATTCGCTTGGCGGCTGGACAGGCTTC 300

```

```

3991 TTCTTAACCTTTCTGTGTGCACTGACCACTGTAACTCATCTCTTCCAGTGAAT 4050
|||||
301 TTCTTAACCTTTCTGTGTGCACTGACCACTGTAACTCATCTCTTCCAGTGAAT 360

```

```

4051 CCTGGAGGCGCTGACCGCTGGTGGCTGTTCAGCTTCCTGCTGGGCGCAGCATTT 4110
|||||
361 CCTGGAGGCGCTGACCGCTGGTGGCTGTTCAGCTTCCTGCTGGGCGCAGCATTT 420

```

```

4111 TTGAGGATTTATCTTAAAGGAGGCTTGCCTCCGTAATTCCTGCTCCCATTTCTC 4170
|||||
421 TTGAGGATTTATCTTAAAGGAGGCTTGCCTCCGTAATTCCTGCTCCCATTTCTC 480

```

```

4171 TCTTGTGAGAGAGATGAGAGCAAGAGTGAAGAAATAGGGGCTGAAGAGCCCA 4230
|||||
481 TCTTGTGAGAGAGATGAGAGCAAGAGTGAAGAAATAGGGGCTGAAGAGCCCA 540

```

```

4231 CTGCCAGATGGCTTCTTCTCTCTCTCTCTTGAACACACGTCCTGGGCTCAG 4290
|||||
541 CTGCCAGATGGCTTCTTCTCTCTCTCTTGAACACACGTCCTGGGCTCAG 600

```

```

4291 GCGTTTCTGAAGTCTCTTCTTGAATGGACAGAGATCAGAGGCTGACATTCCTG 4350
|||||
601 GCGTTTCTGAAGTCTCTTCTTGAATGGACAGAGATCAGAGGCTGACATTCCTG 660

```

```

4351 TGGTCTGAAGTGTGAGAGTGCAGCTCTCTCCCTAGTGAAGAGCAAGCTGTCTCT 4410
|||||
661 TGGTCTGAAGTGTGAGAGTGCAGCTCTCT -CCTAGTGAAGAGCAAGCTGTCTCT 720

```

```

4411 TCGAGGAAACCCAGCGCTGGCGGGAAGTTTACAGCAAGGCGCTGCTTGGGATAT 4470
|||||
720 TCGAGGAAACCCAGCGCTGGCGGGAAGTTTACAGCAAGGCGCTGCTTGGGATAT 779

```

```

4471 TCGTGTGTAATTTACACTTCCCGCGCTCTGTCTGAGAGCCCATCTGTATATCTG 4530
|||||
780 TCGTGTGTAATTTACACTTCCCGCGCTCTGTCTGAGAGCCCATCTGTG -GGTATCTGG 836

```

```

QY 4531 GGTTTTGGACCCCTAATGTCACTTGGCTGTAGACT 4568
|||||
Db 837 GGTTTTGGACCCCTAATGTCACTTGGCTGTAGACT 874

```

## RESULT 6

```

LOCUS AUI25122 803 bp mRNA linear EST 23-OCT-2000
DEFINITION AUI25122 NT2RM4 Homo sapiens cDNA clone NT2RM4001073 5', mRNA
sequence.
ACCESSION AUI25122
VERSION AUI25122.1 GI:10949838
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 803)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
Unpublished (2000)
CONTACT: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```

## REFERENCE

```

AUTHORS
TITLE
JOURNAL
COMMENT

```

## FEATURES

## source

```

1.803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4001073"
/clone_lib="NT2RM4"
/cell_type="teratocarcinoma"
/line="NT2"
/notes="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 177 a 194 c 196 g 233 t 3 others
ORIGIN

```

```

Query Match 12.4%; Score 764.8; DB 9; Length 803;
Best Local Similarity 99.1%; Pred. No. 2.7e-175;
Matches 788; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```

```

QY 3666 TACATTCATACACACTATTATGATTTTAAAAAGAAAGTTTTCAGATTTCCATTTCA 3725
|||||
9 TACATTCATACACACTATTATGATTTTAAAAAGAAAGTTTTCAGATTTCCATTTCA 68

```

```

3726 AGGCTATTATATATATATGCTGTATATATATATATATATATATATATATATATAT 3785
|||||
69 AGGCTATTATATATATATGCTGTATATATATATATATATATATATATATATATAT 128

```

```

3786 ATTTTGGCTGGGAGTGTGAGTTTGGCTTCTAAGGAGGAGCGCGCAGAGCTCTT 3845
|||||
129 ATTTTGGCTGGGAGTGTGAGTTTGGCTTCTAAGGAGGAGCGCGCAGAGCTCTT 188

```

```

3846 TGTTCGTATTTCTGGGAGAGATGGTCTGCGCTTGTGTCACTGCTTATCTTAAAGAT 3905
|||||
189 TGTTCGTATTTCTGGGAGAGATGGTCTGCGCTTGTGTCACTGCTTATCTTAAAGAT 248

```

```

3906 CATCTCCATCTCTCCAGGCGCATCTGTGTGAGCAACCAAGAAAGGATGAACCTTGGCC 3965
|||||
249 CATCTCCATCTCTCCAGGCGCATCTGTGTGAGCAACCAAGAAAGGATGAACCTTGGCC 308

```

OY	3966	CCTTTCGGGCGCCGAGCAAGGCTCCTTCCCTTAACCCCTTTCTGTGCCAGATCAGAACCCTG	4025
Db	309	CTCTTCGGGCGCCGAGCAAGGCTCCTTCCCTTAACCCCTTTCTGTGCCAGATCAGAACCCTG	368
OY	4026	AACTCACATCTCTCTCCAGTGAAATCCCTGGAGGCCCTGACCCCTGGTAGGCTGTTCAG	4085
Db	369	AACTCACATCTCTCTCCAGTGAAATCCCTGGAGGCCCTGACCCCTGGTAGGCTGTTCAG	428
OY	4086	TTCCTGCTCTGGGGCCAGCAGATTTTTGAAGATTATCTTTAAGCCAGGCTTGCCTCGT	4145
Db	429	TTCCTGCTCTGGGGCCAGCAGATTTTTGAAGATTATCTTTAAGCCAGGCTTGCCTCGT	488
OY	4146	ACTTATCCCTGCTCCCTCCATTTCTCTGTTTGAAGAGAAATGAGAGAACGAAGTGA	4205
Db	489	ACTTATCCCTGCTCCCTCCATTTCTCTGTTTGAAGAGAAATGAGAGAACGAAGTGA	548
OY	4206	GAAAGAAATGAGGCGTGAAAGACGCCATCCACAGATGGCTTTCTATCCCTGCTTCTGT	4265
Db	549	GAAAGAAATGAGGCGTGAAAGACGCCATCCACAGATGGCTTTCTATCCCTGCTTCTGT	608
OY	4266	GAAAGAAATGAGGCGTGAGGCGCTCAGGCGCTTCTGAAGTCTCTTTCTTGATTTGACAG	4325
Db	609	GAAAGAAATGAGGCGTGAGGCGCTCAGGCGCTTCTGAAGTCTCTTTCTTGATTTGACAG	668
OY	4326	AGATCAGCAGCGCTGACATCTGCTGTGCTGAAGTGTTTCAGATCCAGCTCCGCTCC	4385
Db	669	AGATCAGCAGCGCTGACATCTGCTGTGCTGAAGTGTTTCAGATCCAGCTCCGCTCC	728
OY	4386	CTAGGTGTAGAGC-AAGCCAGTCTCCTTGAGAGAACCCACCGGCTGG-CGGGAAGTTTT	4443
Db	729	CTAGGTGTAGAGCAAAGCCAGTCTCCTTGAGAGAACCCACCGGCTGGCGGAAGTTTT	788
OY	4444	ACAGCAGAGCGCGCTG 4458                     	
Db	789	ATAGCAGAGCGCGCTG 803	
RESULT	7		
BIS541177		769 bp mRNA linear EST 05-SEP-2001	
LOCUS			
DEFINITION		60323523J1P1 NCL-GAP-Brn67 Homo sapiens cdna clone IMAGE:5309642	
ACCESSION		5' mRNA sequence.	
VERSION		BIS541177	
KEYWORDS		BIS541177.1 GI:15441480	
SOURCE		EST.	
ORGANISM		human.	
REFERENCE		Homo sapiens	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
JOURNAL		1 (bases 1 to 769)	
COMMENT		NH-MGC http://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nhl.gov Tissue Procurement: David N. Louis, M.D. cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LIAK11784 Row: 1 Column: 03 High quality sequence stop: 754.	

```

/lab.host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT      153 a      217 c      216 g      182 t      1 others
ORIGIN

```

BASE COUNT	153 a	217 c	216 g	182 t	1 others
ORIGIN					
Query Match	11.9%	Score 736;	DB 10;	Length 769;	
Best Local Similarity	99.08%	Pred. No. 2.8e-166;			
Matches 761; Conservative	0;	Mismatches 6;	Indels 2;	Gaps 2	



KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
human.	Homo sapiens	1 (bases 1 to 770)	NIH-MGC	http://mgs.nci.nih.gov/	Unpublished (1999)	Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						Tissue Procurement: DCTD/DTF
						CDNA Library Preparation: Ling Hong/Rubin Laboratory
						CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
						DNA Sequencing by: Incyte Genomics, Inc.
						Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov
						Plate: L10CM1944 row: f column: 17
						High quality sequence stop: 770.
FEATURES	source	1..770				Location/Qualifiers
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:5452432"					
	/clone_lib="NIH_MGC_40"					
	/rissue_type="carcinoma, cell line"					
	/lab_host="DH10B (phage-resistant)"					
	/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGACG) library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."					
BASE COUNT	146 a	193 g	225 t			
ORIGIN						
Query Match	11.9%; Score 734.4; DB 10; Length 770;					
Best Local Similarity	99.5%; Pred. No. 7e-168;					
Matches 768; Conservative	0; Mismatches 1; Indels 3; Gaps 3.					
4565	GACCTCCCGAGGTTTGATGTGCTAGAACAAATGGAGGCTGTGATTTGCTGTGTAAGCT	4624				
1	GACTCCCCGAGGTTTGATGTGCTAGAACAAATGGAGGCTGTGATTTGCTGTGTAAGCT	60				
4625	CACATTCACACCTTGGAAATCTAACGGGCACTTCAACCCGAGTTTACCACTTCCACTCCCT	4684				
61	CACATTCACACCTTGGAAATCTAACGGGCACTTCAACCCGAGTTTACCACTTCCACTCCCT	120				
4685	GCTTAAAGATTCCTGTTCCCTGGGGCTGAAGATGAATTAAGCTAATTTTGGTCCACGAGTG	4744				
121	GCTTAAAGATTCCTGTTCCCTGGGGCTGAAGATGAATTAAGCTAATTTTGGTCCACGAGTG	180				
4745	CAGTAGGGGAACCTAGAGAGGTGTGAGTGGCAATTTGTCAAGGATTTAGCCATGACGTGT	4804				
181	CAGTAGGGGAACCTAGAGAGGTGTGAGTGGCAATTTGTCAAGGATTTAGCCATGACGTGT	240				
4805	TTCCTTGAACCCCTACTTCTTGGAAGTGGAGTTGACTCTGGAAGTTTCTTAGCAACTGAAACA	4864				
241	TTCCTTGAACCCCTACTTCTTGGAAGTGGAGTTGACTCTGGAAGTTTCTTAGCAACTGAAACA	300				
4865	AAAGCTCAGGTTTGTCTGCTGATGACATGACATGACGCTTAAAGCCAGTTCCGCTGAGACC	4924				
301	AAAGCTCAGGTTTGTCTGCTGATGACATGACATGACGCTTAAAGCCAGTTCCGCTGAGACC	360				
4925	TTTGCACTCTGCTGCTTCTAATTTCTTGGAATAGCTTCTCCTGACCTGACCTGACACAGT	4984				
361	TTTGCACTCTGCTGCTTCTAATTTCTTGGAATAGCTTCTCCTGACCTGACCTGACACAGT	420				
4985	GGGTCTCTTCAAGTACTGTTTTTGAAGCTGGGGCTCTTTTGTGTAGTCCCAACCACTCT	5044				

D	b		421	GGGTCCTTCAACTAGCTGTGTTGAAGTGCGCTCTTGTGTAGCTCCACCCACCTGT	480
Q	y	5045	AGGCCTAGCTCGCCTTAAGGAACTCTGCCCATTTGGCAAACCGGCGCGCCAG	5104	
D	b	481	AGGGCTAGCTCGCCTTAAGGAACTCTGCCCATTTGGCAAACCGGCGCGCCAG	539	
O	y	5105	GACGTGTGTTCCAAAGGTTGCCCGCCCCAACCCAGATCAGCCTGTAGCTCCCTCT	5164	
D	b	540	GACGTGTGTTCCAAAGGTTGCCCGCCCCAACCCAGATCAGCCTGTAGCTCCCTCT	598	
Q	y	5165	GAGCAGTAGTGTTATGTATGTTCCACAGAGTGGGGGTCAGACGCCCTTCTCAGAACTTCTA	5224	
D	b	599	GAGCAGTAGTGTTATGTATGTTCCACAGAGTGGGGGTCAGACGCCCTTCTCAGAACTTCTA	658	
Q	y	5225	GTTGCCCTTACCTGACTCTCTGACTTGTATTCTTTTAGCAGTAGCCTTCTTCCCTCGGG	5284	
D	b	659	GTTGCCCTTACCTGACTCTCTGACTTGTATTCTTTTAGCAGTAGCCTTCTTCCCTCGGG	718	
Q	y	5285	GAGCC-AAAGAGTGTGTGTGGGGCGCTATPATYTGCGCTATTTCATCTG	5335	
D	b	719	GAGCCAAAAGAGTGTGTGTGGGGCGCTATPATYTGCGCTATTTCATCTG	770	
R	E	S	RESULT	9	
L	O	C	AUI40433	743 bp	mRNA linear EST 25-OCT-2000
D	E	F	INIATION	AUI40433 PLACE3 Homo sapiens cDNA clone PLACE3000047 5', mRNA	
A	C	C	SSION	AUI40433	sequence.
V	E	R	SION	AUI40433	
K	E	N	WORDS	AUI40433.1 GI:11001954	EST.
S	O	U	RCE	Homo sapiens	
O	R	G	NISM	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
R	E	F	ERENCE	1 (bases 1 to 743)	
A	U	T	HORS	Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S., Masuno,Y. and Isogai,T.	
T	I	T	LE	HRI human CDNA project (Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S., Masuno,Y., Isogai,T.)	
J	O	R	NAL	Unpublished (2000)	
C	O	M	MENT	Contact: Takao Isogai	
				Genomics Laboratory	
				Helix Research Institute	
				1532-3 Yana, Kisarazu, Chiba 292-0812, Japan	
				Tel.: 81-438-52-3951	
				Fax: 81-438-52-3952	
				Email: genomics@hri.co.jp	
				HRI human CDNA Project: 5'- & 3'-end one pass sequencing: Helix	
				Research Institute; cDNA library construction: Department of	
				Virology, Institute of Medical Science, University of Tokyo, and	
				Helix Research Institute.	
F	E	A	TURES	Location/Qualifiers	
S	O	U	RCE	1..743	
				/organism="Homo sapiens"	
				/db_xref="taxon:9606"	
				/clone="PLACE3000047"	
				/clone_lib="PLACE3"	
				/tissue_type="Placenta"	
				/note="Vector: pHE18SFL3"	
B	A	S	E COUNT	154 a	204 c
O	R	I	GIN	154 g	201 g
				181 t	3 others
Q	U	E	R	Y Match	11.8% Score 725.8 DB 9: Length 743:
B	E	S	T Local Similarity	99.2% Pred. No. 8.6e-166:	
M	A	T	C	hes 738: Conservative	0: Mismatches 5: Indels 1: Gaps
Q	U	E	R	Y 2125	AAAGCGCGTTTCTGCGGAGTACTTCTCCACTGCTGACAGTATTTTAAATACGTCG 2184
D	B	1	AAAGCGCGTTTCTGCGGAGTACTTCTCCACTGCTGACAGTATTTTAAATACGTCG	60	

QY 2185 TGGCTATGATTTAGAGATCGCATCCAGATCCAGCCGCTCCATCCGCGAAGAGCGCGAG 2244  
|||||  
Db 61 TGGCTATGATTTAGAGATCGCATCCAGATCCAGCCGCTCCATCCGCGAAGAGCGCGAG 120  
QY 2245 CTGAGATCCCGCCAGCGCTTCAGGCAAGGCGGTGGAAGTACCCCAACATGCTGACAAAG 2304  
|||||  
Db 121 CTGAGATCCCGCCAGCGCTTCAGGCAAGGCGGTGGAAGTACCCCAACATGCTGACAAAG 180  
QY 2305 GGGACCTGGACCTGGGCGCAGTGGAGAGAGGAGCCGCCCTACGCTCACCAGATACCTCC 2364  
|||||  
Db 181 GGGACCTGGACCTGGGCGCAGTGGAGAGAGGAGCCGCCCTACGCTCACCAGATACCTCC 240  
QY 2365 GCAATTTACTCTGACTGCTCACTGCTTTTACCGAGCTGGAAGTGTCTCCAGGCGCTT 2424  
|||||  
Db 241 GCAATTTACTCTGACTGCTCACTGCTTTTACCGAGCTGGAAGTGTCTCCAGGCGCTT 300  
QY 2425 TCCATCTTTTCCCGCCACACATCCCTTTGACCTGTGTGCTCCAGAGGCCAAATCTTGG 2484  
|||||  
Db 301 TCCATCTTTTCCCGCCACACATCCCTTTGACCTGTGTGCTCCAGAGGCCAAATCTTGG 360  
QY 2485 GCTAGGAGCTGGGCTTGGGCTTATGTTGTGTGGAGCTGGGGGTTCTGGCAAGATTTTC 2544  
|||||  
Db 361 GCTAGGAGCTGGGCTTGGGCTTATGTTGTGTGGAGCTGGGGGTTCTGGCAAGATTTTC 420  
QY 2545 AGGAGGCGCTTGGAGGCAATGTTCCCGAGAGAGAGTTCAGCGTGTGTGTTGACTTATG 2604  
|||||  
Db 421 AGGAGGCGCTTGGAGGCAATGTTCCCGAGAGAGAGTTCAGCGTGTGTGTTGACTTATG 480  
QY 2605 AGCGGGAGGAAGTCTTATGACTTCTTTAGAGAGGCTGAATGGCTCCCTTACCTGAACA 2664  
|||||  
Db 481 AGCGGGAGGAAGTCTTATGACTTCTTTAGAGAGGCTGAATGGCTCCCTTACCTGAACA 540  
QY 2665 AGGCTGCTGCTGTGGAAATCTCCCAAGCTGGCATCAGAGGACTTGTGGGCTGACA 2724  
|||||  
Db 541 AGGCTGCTGCTGTGGAAATCTCCCAAGCTGGCATCAGAGGACTTGTGGGCTGACA 600  
QY 2725 TTGGCGTTCCCATCAGTGTGTGCTGCTAGTGAAGACAGTTTGAACACGATTTCTTAC 2784  
|||||  
Db 601 TTGGCGTTCCCATCAGTGTGTGCTGCTAGTGAAGACAGTTTGAACACGATTTCTTAC 660  
QY 2785 CCGGAAATGAATTTGAGACAGAGGCCATCTCTCCATTTGATGAGATGCTACCTCCGC 2844  
|||||  
Db 661 CCGGAAATGAATTTGAGACAGAGGCCAT-CTGTCTATTTGATGAGATGCTACCTCCGC 719  
QY 2845 ATGACGAAATCATGTTTGGGTTCC 2868  
|||||  
Db 720 ATGACNAAATCATGTTTGGTTCC 743

RESULT 10  
B1196587 763 bp mRNA 1linear EST 10-JUL-2001  
LOCUS 602755556f1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4891261 5',  
DEFINITION mRNA sequence.  
ACCESSION B1196587  
VERSION B1196587.1 GI:14651607  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 763)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

Found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCMI769 row: p column: 14  
High quality sequence start: 4  
High quality sequence stop: 761.  
FEATURES  
source  
1..763  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4891261"  
/clone\_lib="NIH\_MGC\_19"  
/tissue\_type="neuroblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: brain; Vector: pOT87; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGGCAGAG(G) Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using Zap-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC library."  
BASE COUNT 144 a 230 c 218 g 171 t  
ORIGIN  
Query Match 11.6%; Score 715; DB 10; Length 763;  
Best Local Similarity 97.6%; Pred. No. 3.8e-163;  
Matches 747; Conservative 0; Mismatches 15; Indels 3; Gaps 2;  
QY 1761 AAGCTGATCAGGCTCGTGGTGAATTCACCTGCAAAAAACAGCCCAACAGCTCCG 1820  
|||||  
Db 2 AAGCTGATCAGGCTCGTGGTGAATTCGCAAAAAACAGCCCAACAGCTCCG 61  
QY 1821 ACTGAGTGGGACATGTGTGAGAGCGGAGAGCGCTTGAATGCTGAAGCTCTCCAGC 1880  
|||||  
Db 62 ACTGAGTGGGACATGTGTGAGAGCGGAGAGCGCTTGAATGCTGAAGCTCTCCAGC 121  
QY 1881 TTGGCCCTCATATTACCCCGGGAGCCCTGCGTGTGTTATTTCTCTGGGTGCAACA 1940  
|||||  
Db 122 TTGGCCCTCATATTACCCCGGGAGCCCTGCGTGTGTTATTTCTCTGGGTGCAACA 181  
QY 1941 CGGCTCTTCAGAGCCCGTGAAGTGGTGGTCCCGGTGGTGGTGGGGGAGAGGTCCAG 2000  
|||||  
Db 182 CGGCTCTTCAGAGCCCGTGAAGTGGTGGTCCCGGTGGTGGTGGGGGAGAGGTCCAG 241  
QY 2001 CTTCCCTACAGGACATGCTGAGTGAAGAGAGCGCCCTGTGGTGGCCAAAGCTCGT 2060  
|||||  
Db 242 CTTCCCTACAGGACATGCTGAGTGAAGAGAGCGCCCTGTGGTGGCCAAAGCTCGT 301  
QY 2061 GTTACCGAGGTTCATTCTCGCTCAGAGAGCTCTCCGATAGTACCTCTGGCTATGAG 2120  
|||||  
Db 302 GTTACCGAGGTTCATTCTCGCTCAGAGAGCTCTCCGATAGTACCTCTGGCTATGAG 361  
QY 2121 CGGCAAGGCGCTTCTCTGGAGAGCTTCTTCACAGTGTGACGTATTTTATATAC 2180  
|||||  
Db 362 CGGCAAGGCGCGTCTTCTCTGGAGAGCTTCTTCACAGTGTGACGTATTTTATATAC 421  
QY 2181 GTTCTGGCTATGATTAAGATCGCATCCAGATCCAGCCGCTCCATCCGGAAGAGGG 2240  
|||||  
Db 422 GTTCTGGCTATGATTAAGATCGCATCCAGATCCAGCCGCTCCATCCGGAAGAGGG 481  
QY 2241 GGAGCTGAATCCCGCCAGCGCTTCAGGCAAGGCGGTGAGATGACCCCAATGCTGAC 2300  
|||||  
Db 482 GGAGCTGAATCCCGCCAGCGCTTCAGGCAAGGCGGTGAGATGACCCCAATGCTGAC 541  
QY 2301 AACGGGAGCTGGACCTGGGGCCAGTGGAGAGGAGCGCCCTAGGCTCACCAGATAC 2360  
|||||  
Db 542 AACGGGAGCTGGACCTGGGGCCAGTGGAGAGGAGCGCCCTAGGCTCACCAGATAC 601  
QY 2361 CTCGCAATTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2420  
|||||  
Db 602 CTCGCAATTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661  
QY 2421 CTTTTCATCTTTTCCCGCAGCTCCTTTGACCTGTGTGCTGCTGCTGCTGCTGCTGCTG 2480

Db	652	CCTTTCCATCTTTTCCCA-AGTCCCTTTGACCCTGTGTGGC--TCAGAGCAAAATTC	718
Oy	2481	TTGGGCTCAGGACTGCTTTGCGCCATTTGGTGTGAGCTGGC	2525
Db	719	TTGGGCTCAGGACTGCTTTGCGCACTATAGTGTGAGCTGGG	763
RESULT	11		
LOCUS	AU132259	734 bp	linear
DEFINITION	AU132259 NT2RP3 Homo sapiens cDNA clone NT2RP3004092 5', mRNA		
ACCESSION	AU132259		
VERSION	AU132259.1	GI:10992613	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogami,T.		
TITLE	HRI human cDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1. 734 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RP3004092" /clone_11b="NT2RP3" /clone_type="teratocarcinoma" /cell_line="NT2" /note="vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"		
FEATURES			
SOURCE			
CASE COUNT	113 a	193 c	197 g
ORIGIN			168 t
Query Match	11.4%;	Score 706.4;	DB 9;
Best Local Similarity	99.2%;	Pred. No. 4,6e-161;	
Matches 729;	Conservative 0;	Mismatches 4;	Indels 2;
		Gaps 2;	
Oy	3032	TTCTTATGTGATGCCCCAGCGCCATCCGGAGATGGTGATGAATACATCACTGTGAGGA	3091
Db	1	TTCTTATGTGATGCCCCAGCGCCATCCGGAGATGGTGATGAATACATCACTGTGAGGA	60
Oy	3092	CATTGCGCATGAACCTTCTTGTCTCCACATCACTGGGAAGCCCCCATCAAGGTGACCTC	3151
Db	61	CATTGCGCATGAACCTTCTTGTCTCCACATCACTGGGAAGCCCCCATCAAGGTGACCTC	120
Oy	3152	ACGGTGAGCATTCGATGCGCCAGAGATGCCCTGAGGCCCTGTCTCATGATGATCCCACTT	3211
Db	121	ACGGTGAGCATTCGATGCGCCAGAGATGCCCTGAGGCCCTGTCTCATGATGATCCCACTT	180
Oy	3212	CCAGCAGCGGCACAAGTGATCAACTTCTTCTGTAAGGTGACGCGCTACATGCCCCCTCT	3271
Db	181	CCAGCAGCGGCACAAGTGATCAACTTCTTCTGTAAGGTGACGCGCTACATGCCCCCTCT	240
Oy	3272	GTACACGCACTTCAGGCTGTGATTTGTGCTTTCAAGACAGCGCTGCCCATGACAAAGC	3331

Db	241	GTACACGAGATTTCAGGGGTGGATTTCTGTCTCCTTCAAGACACAGCCCTGCCCATATGACAAAGC	300
Oy	3332	CAGTGCTTCAAGTTCATCTAGGGGCACCAGCGCTCGGGGAAGAGATATGACAGAGG	3391
Db	301	CAAGGCTTCAAGTTCATCTAGGGGCACGCGCTCGGGGAAGAGATATGACAGAGG	360
Oy	3392	AGGAAGATGGCTCCCAAGGTTCTTAGGCAATTGCAGACCTTGGGCACATCTGTGGTGG	3451
Db	361	AGGAAGATGGCTCCCAAGGTTCTTAGGCAATTGCAGACCTTGGGCACATCTGTGGTGG	420
Oy	3452	TGGCCACAGAGCCCTGTCTGGAAAGGGGCAGCAGAGAGATGGAAAGGAACCCGTGCTTTA	3511
Db	421	TGGCCACAGAGCCCTGTCTGGAAAGGGGCAGCAGAGAGATGGAAAGGAACCCGTGCTTTA	480
Oy	3512	TCTTGAAGTCAAGCCACACTGGGCTCTGGAGCCCTGGGGGAGTCCCCGGGTTTCCCACAC	3571
Db	481	TCTTGAAGTCAAGCCACACTGGGCTCTGGAGCCCTGGGGGAGTCCCCGGGTTTCCCACAC	540
Oy	3572	AGGGCACAGACGTGATAGTTCACATGAGAGACGTGGGACACTGTGACAGTCACTCACAC	3631
Db	541	AGGGCACAGACGTGATAGTTCACATGAGAGACGTGGGACACTGTGACAGTCACTCACAC	600
Oy	3632	GTTCTAGCCGCCAGAGAC-AGCTGGTTCGTGGTTTTACATTCATTAACAACACTATTATGAT	3690
Db	601	GTTCTAGCCGCCAGAGACAGCTGGTTCGTGGTTTTACATTCATTAACAACACTATTATGAT	660
Oy	3691	TATTTAAAAAGAAGAAAGTTTCAGATTGCCATTCAGGCTTATTTATATATATGTGTG	3750
Db	661	TATTTAAAAAGAAGAAAGTTTCAGATTGCCATTCAGGCTTATTTATATATG-GANG	719
Oy	3751	TATATAATACATGC 3765 	
Db	720	NATATAAATACATGC 734 	
<b>RESULT 12</b>			
AU133145			
LOCUS	AU133145	768 bp	mRNA linear EST 24-OCT-2000
DEFINITION	AU133145 NT2RP4 Homo sapiens cDNA clone NT2RP4001395 5', mRNA sequence.		
ACCESSION	AU133145		
VERSION	AU133145.1	GI:10993684	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 768) Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Salto,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuno,Y., and Isogai,T. HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Salto,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuno,Y., Isogai,T.) Unpublished (2000)		
JOURNAL	Contact: Takao Isogai		
COMMENT	Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel.: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomese@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..768 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RP4001395" /clone_1kb="NT2RP4" /cell_type="teratocarcinoma"		
FEATURES	source		

/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"  
BASE COUNT 163 a 200 c 202 g 199 t 4 others  
ORIGIN

Query Match 11.4% Score 706; DB 9; Length 768;  
Best Local Similarity 97.9%; Pred. No. 5.9e-161;  
Matches 735; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 2477 ATCTTGGGCTCAGGAGACTGGCTTCCGCTATTGTTGGTGGAGCTGGGGTTCGGCAA 2536  
DB 1 ATCTTGGGCTCAGGAGACTGGCTTCCGCTATTGTTGGTGGAGCTGGGGTTCGGCAA 60  
QY 2537 GGAATTTAGGACGCGCTTGGAGCAATGTTCCCGAGAGCAGTTACAGGTGTGATGT 2596  
DB 61 GGAATTTAGGACGCGCTTGGAGCAATGTTCCCGAGAGCAGTTACAGGTGTGATGT 120  
QY 2597 GACTTATGAGGGGAGGAGGCTTATGAACTTTAGAGAGGCTGAAATGCTCCCTTA 2656  
DB 121 GACTTATGAGGGGAGGAGGCTTATGAACTTTAGAGAGGCTGAAATGCTCCCTTA 180  
QY 2657 CCTGAAACAGGCTGCTGTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTGTG 2716  
DB 181 CCTGAAACAGGCTGCTGTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTGTG 240  
QY 2717 GCTGACATTTGGCGTTCCCATCATGTGTGCTCCGACTGAGAAACAGTTTGAACACCG 2776  
DB 241 GCTGACATTTGGCGTTCCCATCATGTGTGCTCCGACTGAGAAACAGTTTGAACACCG 300  
QY 2777 ATCTTCCCTCGAATGAATGAGACAGGCAATCCTGCTCATGTATGACATGCTCA 2836  
DB 301 ATCTTCCCTCGAATGAATGAGACAGGCAATCCTGCTCATGTATGACATGCTCA 360  
QY 2837 CCTCCGCGATGACAAATCATGTTGGTTCGGGTGTGAGAGAAAGCTCGGACCGCAT 2896  
DB 361 CCTCCGCGATGACAAATCATGTTGGTTCGGGTGTGAGAGAAAGCTCGGACCGCAT 420  
QY 2897 CGTGGGCTTCCCTGGCGTTACACAGCATGGGACATCCCATCATGCTTGGCTCAAA 2956  
DB 421 CGTGGGCTTCCCTGGCGTTACACAGCATGGGACATCCCATCATGCTTGGCTCAAA 480  
QY 2957 CTCCACATGCTGAGAGGCTCATGAGGCTGACAGGCTGCTGCTTTCACAAGTA 3016  
DB 481 CTCCACATGCTGAGAGGCTCATGAGGCTGACAGGCTGCTGCTTTCACAAGTA 540  
QY 3017 TTATGCTACCTGATTTCTATGTATGATGCCAGGCGCATCGGAGCATGTGATGATA 3076  
DB 541 TTATGCTACCTGATTTCTATGTATGATGCCAGGCGCATCGGAGCATGTGATGATA 600  
QY 3077 CATCAACTGTGAGACATTTGCCATGAACCTTCTTCTCCACATCAGCTCGGAAGCCCC 3136  
DB 601 CATCAACTGTGAGACATTTGCCATGAACCTTCTTCTCC - ACATCTCGGAAGCCCC 658  
QY 3137 CATCAAGGTGACCTCAAGGAGGACATTCGATGCCAGGATGCCCTAGAGCCC - TGCTC 3195  
DB 659 CATCAAGGTGACCTCAAGGAGGACATTTGATGATGCCCAAGATGCCCTTAGCTC 718  
QY 3196 ATGATGACTCCACTTCCAGCAGCGGACAA 3226  
DB 719 ATGATGACTCCACTTCCAGCAGCGGACAA 749

RESULT 13  
BG115714  
LOCUS BG115714 1000 bp mRNA linear EST 30-JAN-2001  
DEFINITION 602317188F1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:441772.5',  
ACCESSION BG115714  
VERSION BG115714.1 GI:12609220  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1000)  
AUTHORS NIH-MGC  
TITLE NIH-MGC  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LRAM10150 row: 0 column: 21  
High quality sequence stop: 704.

FEATURES  
Source Location/Qualifiers  
1..1000

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:441772"  
/clone\_lib="NIH\_MGC\_88"  
/tissue\_type="duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: small intestine; Vector: pCMV-Sport6;  
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
oligo-dr primed. Average insert size 1.767 kb. Library  
enriched for full-length clones and constructed by Life  
Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 218 a 260 c 294 g 227 t 1 others  
ORIGIN

Query Match 10.9%; Score 672.4; DB 10; Length 1000;  
Best Local Similarity 93.5%; Pred. No. 1e-152;  
Matches 747; Conservative 0; Mismatches 42; Indels 10; Gaps 4;

QY 4399 AGCCAGTGTCTTCGAGAAACCCACCGGCTGGCGGAGTTTACAGCAAGCGGCTG 4458  
DB 1 AGCCAGTGTCTTCGAGAAACCCACCGGCTGGCGGAGTTTACAGCAAGCGGCTG 60  
QY 4459 CTTGGGATTAATTCCTTGTGAAATTCACCTTCCCGGCTGTGCTGAGAGCCCATTC 4518  
DB 61 CTTGGGATTAATTCCTTGTGAAATTCACCTTCCCGGCTGTGCTGAGAGCCCATTC 120  
QY 4519 TGTGTTATCTGTGGTTTGTGAGCCCAATGATGAGTGTGCTGAGAGCTCCCGAGTT 4578  
DB 121 TGTGTTATCTGTGGTTTGTGAGCCCAATGATGAGTGTGCTGAGAGCTCCCGAGTT 180  
QY 4579 TGTGTTATCTGTGAGAACTGAGAGGCTGATTTGCTGTGATGATCAATCCAGCCTTG 4638  
DB 181 TGTGTTATCTGTGAGAACTGAGAGGCTGATTTGCTGTGATGATCAATCCAGCCTTG 240  
QY 4639 GAATCTAACGGGATTCACAAACCGAGTTACACTTTCACCTCCCTGATTAGATTCTGT 4698  
DB 241 GAATCTAACGGGATTCACAAACCGAGTTACACTTTCACCTCCCTGATTAGATTCTGT 300  
QY 4699 TCCTTGGGCTGAAACCTGAATTAAGCTAATTTTGGGTACAGGTGCGCAGTAGGGAACCT 4758  
DB 301 TCCTTGGGCTGAAACCTGAATTAAGCTAATTTTGGGTACAGGTGCGCAGTAGGGAACCT 360  
QY 4759 AGGAGGCTGAGTGGCATTTGACAGGAGTTTGGCCATGAGCTGTTCTGAACCTTAC 4818  
DB 361 AGGAGGCTGAGTGGCATTTGACAGGAGTTTGGCCATGAGCTGTTCTGAACCTTAC 420  
QY 4819 TTCTCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4878  
DB 421 TTCTCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 480  
QY 4879 TCCTGCTCATGACATGCTTAAAGCCAGTTCCGCTTCCCTAGACCTTGACCTGTGCTC 4938

Db 481 TCCTGTCATGACAGACCCCTTAAGCCAGTTCCTCTTCCTTAGACCTTGGCATCTGCTGC 540

Qy 4939 TTCTATTCTTGGAAATACGTTCTCTCTGACCTGCTACCAAGTGGTCTCTTCAAG 4998  
|||||

Db 541 TTCTATTCTTGGAAATACGTTCTCTCTGACCTGCTACCAAGTGGTCTCTTCAAG 600  
|||||

Qy 4999 TACTGTTTGAAGCTGGGCTCTTTTGTGTAGCTCCACCCAGCTGTAGGCTAGCTGGC 5058  
|||||

Db 601 TACTG-TTTGAACCTGGGCTCTTTTGTGTAGCTCCACCAACCTGTAGGCTAGCTGG- 658  
|||||

Qy 5059 TTTAAGGAATCTCCCATTTGGCAACCGGAGCGCGCGGCGGAGCTGTGTCCAA 5118  
|||||

Db 659 TTTAAGGAATCTCTCCCATTTGGCAACCGGAGCGCGG---CAGAGCTGGGTTCACAA 714  
|||||

Qy 5119 AGGTTCCCGCCCGCCCAACCCAGCATCAGCTGTAGCTCCCTGCTGAGGAGCATGTGTT 5178  
|||||

Db 715 AGGTTCCCGCCCGCCCAACCCAGCATC---GGCGGGAAGTCCGGTGTAGGAGGAGGGGAT 770  
|||||

Db 5179 ATGTTCCAGCAGCTGGGGG 5197  
|||||

Db 771 ATGTTCCAGCAGCAAGCGGG 789  
|||||

RESULT 14

AU133571 755 bp mRNA linear EST 24-OCT-2000

LOCUS AU133571 OVARC1 Homo sapiens cDNA clone OVARC1000181 5', mRNA

DEFINITION sequence.

ACCESSION AU133571

VERSION AU133571.1 GI:10994110

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 755)

AUTHORS Oda,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.

TITLE HRI human cDNA project

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source

1..755

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone.lib="OVARC1000181"

/clone.lib="OVARC1"

/tissue\_type="ovary, tumor tissue"

/note="Vector: pME18SFL3"

BASE COUNT 151 a 181 c 201 g 219 t 3 others

ORIGIN

Query Match 10.8%; Score 666.2; DB 9; Length 755;  
Best Local Similarity 97.1%; Pred. No. 3e-151; Indels 5; Gaps 3;  
Matches 709; Conservative 0; Mismatches 16;

Qy 4109 TTTTGAAGATTTATCTTTAGCCAGGCTGCTGCTACTTATCCCTGCTCCATTTC 4168  
|||||

Db 1 TTTTGAAGATTTATCTTTAGCCAGGCTGCTGCTACTTATCCCTGCTCCATTTC 60  
|||||

Qy 4169 TCTCTTTTGAAGAGATGAGCAACCAAGAGTGAAGAAATAGGGCTGAAGAGCG 4228  
|||||

Db 61 TCTCTGTTTGAAGAGAAAGAGAGCAAGAGTGAAGAAATAGGGCTGAAGAGCGC 120  
|||||

Qy 4229 CACTCCAGATGGCTCTTTTATCTGCTCTCTCTGTTGAACAACAGTGTGGGCTTC 4288  
|||||

Db 121 CACTCCAGATGGCTCTTTTATCTGCTCTCTCTGTTGAACAACAGTGTGGGCTTC 180  
|||||

Qy 4289 AGCGTTTCTGAAGTGTCTTTTCTTGTGATTTGACAGAGATGACAGCGTGCATCTGC 4348  
|||||

Db 181 AGCGTTTCTGAAGTGTCTTTTCTTGTGATTTGACAGAGATGACAGCGTGCATCTGC 240  
|||||

Qy 4349 TGTGTCGTGAAGTGTGTTTGTGAGTCAAGCTCTCTCTAGTGTAGAGCAAGCAAGTGC 4408  
|||||

Db 241 TGTGTCGTGAAGTGTGTTTGTGAGTCAAGCTCTCTCTAGTGTAGAGCAAGCAAGTGC 300  
|||||

Qy 4409 CTTGAGAGAACCCAGCCGCGTGGCGGGAAGTTTTCAGAGAGCGGCTCTTGGATA 4468  
|||||

Db 301 CTTGAGAGAACCCAGCCGCGTGGCGGGAAGTTTTCAGAGAGCGGCTCTTGGATA 360  
|||||

Qy 4469 ATTCTTGTGTAATTCACCTTCCCGCCGCTCTGCTGTGAGGCCCATCTGTATCT 4528  
|||||

Db 361 ATTCTTGTGTAATTCACCTTCCCGCCGCTCTGCTGTGAGGCCCATCTGTATCT 420  
|||||

Qy 4529 GTGTTTGTGAGCCCTTAATGTACGCTGTGAGTCCCGGAGGTTTGTATGTC 4588  
|||||

Db 421 GTGTTTGTGAGCCCTTAATGTACGCTGTGAGTCCCGGAGGTTTGTATGTC 480  
|||||

Qy 4589 TAGAACATGGAGGCGTGTATTTGCTGTAGTCAATCCAGCTTGGAAATCTAACG 4648  
|||||

Db 481 TAGAACATGGAGGCGTGTATTTGCTGTAGTCAATCCAGCTTGGAAATCTAACG 540  
|||||

Qy 4649 GGCATTCACAAACCCAGATTCACATTCACCTCCCTTGTAGATTCGTTCCTGGCT 4708  
|||||

Db 541 GGCATTCACAAACCCAGATTCACATTCACCTCCCTTGTAGATTCGTTCCTGGCT 600  
|||||

Qy 4709 GAACATGAATTAAGCTTAATTTTGGTCAAGTGGGAGTA-GGGGAAGCTAGAGGCTG 4767  
|||||

Db 601 GAACATGAATTAAGCTTAATTTTGGTCAAGTGGGAGTA-GGGGAAGCTAGAGGCTG 660  
|||||

Qy 4768 TGAAGTG-CAATTTGTCAGGATTTAGCCCATGACGTG--TTTCTGAACCTACTTCT 4823  
|||||

Db 661 TGAAGTGCGATTTGGCAAGGATTTAGCCCATGACGTGTTTCTTGAACCTACTTCT 720  
|||||

Qy 4824 GGAAGTGGAG 4833  
|||||

Db 721 TGGAAAGNGG 730  
|||||

RESULT 15

BF982138 951 bp mRNA linear EST 23-JAN-2001

LOCUS BF982138

DEFINITION 602308950F1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4400196 5', mRNA sequence.

ACCESSION BF982138

VERSION BF982138.1 GI:12384950

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 951)

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 18:50:09 ; Search time 108.24 Seconds  
(without alignments)  
14006.375 Million cell updates/sec

Title: US-09-809-920-3

Perfect score: 6172  
Sequence: 1 GCGGGGTCCTGAGCTGGA.....ATTAATGAGACTTAACTC 6172

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/1na/PCITUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/1na/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199	3.2	3003	4 US-08-915-337-1	Sequence 1, Appl
2	57.2	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl
3	46	0.7	2241	2 US-08-838-2198-20	Sequence 20, Appl
4	46	0.7	2241	3 US-09-233-336A-20	Sequence 20, Appl
5	46	0.7	2241	4 US-09-233-752A-20	Sequence 20, Appl
6	46	0.7	2241	5 US-09-402-036-20	Sequence 20, Appl
7	46	0.7	2370	2 US-08-838-2198-19	Sequence 19, Appl
8	46	0.7	2370	3 US-09-233-336A-19	Sequence 19, Appl
9	46	0.7	2370	4 US-09-233-752A-19	Sequence 19, Appl
10	46	0.7	2370	5 US-09-402-036-19	Sequence 19, Appl
11	46	0.7	2403	1 US-08-471-044-30	Sequence 30, Appl
12	46	0.7	2403	2 US-08-471-044-30	Sequence 30, Appl
13	46	0.7	2403	3 US-08-463-483A-30	Sequence 30, Appl
14	46	0.7	2403	4 US-08-471-046A-30	Sequence 30, Appl
15	46	0.7	2403	5 US-08-470-566B-30	Sequence 30, Appl
16	46	0.7	2403	6 US-08-838-2198-7	Sequence 7, Appl
17	46	0.7	2403	7 US-08-838-2198-7	Sequence 7, Appl
18	46	0.7	2403	8 US-08-469-334-30	Sequence 30, Appl
19	46	0.7	2403	9 US-09-300-529-30	Sequence 30, Appl
20	46	0.7	2403	10 US-09-233-336A-7	Sequence 7, Appl
21	46	0.7	2403	11 US-09-233-752A-7	Sequence 7, Appl
22	46	0.7	1435	4 US-09-402-036-7	Sequence 7, Appl
23	44.6	0.7	1890	6 US-09-153-804-4	Sequence 4, Appl
24	43.8	0.7	16442	3 US-08-781-891-208	Sequence 208, App
25	43.6	0.7	1361	4 US-09-428-583-3	Sequence 3, Appl
26	43.6	0.7	1700	2 US-08-839-581A-1	Sequence 1, Appl
27	43.6	0.7	1700	4 US-09-023-591A-1	Sequence 1, Appl

C 28	43.6	0.7	10763	1 US-08-761-258-1	Sequence 1, Appl
C 29	43.6	0.7	10763	2 US-08-977-306-1	Sequence 1, Appl
C 30	43.4	0.7	1368	3 US-08-874-563-5	Sequence 5, Appl
C 31	43.4	0.7	1368	4 US-08-577-483-14	Sequence 14, Appl
C 32	43.4	0.7	4253	3 US-08-577-483-7	Sequence 7, Appl
C 33	43.2	0.7	4254	2 US-08-443-639-7	Sequence 7, Appl
C 34	42.4	0.7	3709	4 US-08-541-782-7	Sequence 7, Appl
C 35	42.2	0.7	3572	2 US-08-713-815A-2	Sequence 2, Appl
C 36	41.6	0.7	43280	2 US-08-804-227C-1	Sequence 1, Appl
C 37	41.2	0.7	80246	4 US-09-078-294-4	Sequence 4, Appl
C 38	40.4	0.7	51952	3 US-08-947-823-1	Sequence 1, Appl
C 39	40.2	0.7	1048	4 US-09-153-804-1	Sequence 1, Appl
C 40	40	0.6	1767	1 US-07-903-466-2	Sequence 2, Appl
C 41	40	0.6	1767	5 PCT-US93-05794-2	Sequence 2, Appl
C 42	40	0.6	3018	1 US-07-903-466-1	Sequence 1, Appl
C 43	40	0.6	3018	5 PCT-US93-05794-1	Sequence 1, Appl
C 44	40	0.6	7218	1 US-08-232-463-14	Sequence 14, Appl
C 45	39.6	0.6	939	1 US-08-285-440-10	Sequence 10, Appl

# ALIGNMENTS

RESULT 1  
US-08-915-337-1  
Sequence 1, Application US/08915337  
Patent No. 6287802  
GENERAL INFORMATION:  
APPLICANT: Deng, Fan & Xia  
TITLE OF INVENTION: EXT2 Gene  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Smtlinkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTED Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915.337  
FILING DATE: August 21, 1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: China - 96121928.9  
FILING DATE: October 21, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William T. Han  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: ATG50019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 270-5219  
TELEFAX: (610) 270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3003 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
HYPOTHETICAL: NO  
AMTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-915-337-1  
Query Match 3.2% Score 199; DB 4; Length 3003;



Best Local Similarity 57.8%; Pred. No. 3.9e-45;  
Matches 396; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

```
OY 2668 TCGTGGGATGATGATTCCTCCCAAGCTGCGATCAGAGAGACCTTCTGTGGCTGACATTG 2727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1653 TTGTCTCTGGGAATATCAAGATAAAACCTCCAGAAAGATTCTCTGTGCCCAAAATCC 1712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2728 GCGTCCCATCATGATGTCCTCCGATGAGAAAGACATGTTGAACCCGATTCCTACCC 2787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1713 GGGTTCATTAAAGTTTGAGAGACTGCTGAAACAAAGTTAAAGTAAACCGTTCTCCCTT 1772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2788 GGAATGAATGAGACAGAGAGCCATCTCTCATTTATGATGATGATGATGATGATGATG 2844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1773 ATGATGAATTCAGACAGAGAGCTGTCTGCGCATGATGATGATGATGATGATGATGATG 1832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2845 ATGAGCAATTCATGTTGGGTTCCGGGTGTGAGAGAACTGGGAGCCCATCTGGGCT 2904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1833 CTGACAGCTGCAATTTGGTATGAGGTCTGGCGGGAATTTCTCCGCGGTGGTGGT 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2905 TCCCTGGCCGTTACCAAGCATGAGATCCCATCATGCTCCGCTCTCACTCACTCACT 2964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1893 ACCCGGATCGTCTGATCTCTGGACCATGATGATGATGATGATGATGATGATGATGATG 1952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2965 ACTCTGTGAGCTGTCCATGATGTCGACAGTGTGCTCTTTCACAGATTTATGCT 3024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1953 GGACGAATGAAGTTCATGATGTCATCTGGGAGCTTTTATACAGATTTTAAAT 2012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3025 ACCCTATTCTTATGATGATGCTCCAGCCATCCGAGCATGATGATGATGATGATGATGAT 3084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2013 ACCCTATTCTTATGATGATGCTCCAGCCATCCGAGCATGATGATGATGATGATGATGATG 2072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3085 GTGAGACATTTCCATGATGATGCTCTGTCATCATGATGATGATGATGATGATGATGATG 3144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2073 GTGAGATATTTCCATGATGATGCTCTGTCATCATGATGATGATGATGATGATGATGATG 2132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3145 TGACCTACGCTGATGATGATGCTCCAGCCATCCGAGCATGATGATGATGATGATGATG 3198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2133 TAACCCCGCAAGAAATTTCAAGTCTGATGATGATGATGATGATGATGATGATGATGATG 2192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3199 ATGATCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2193 ACCAATCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3259 ACATGCCCTCTCTACACGATGATGATGATGATGATGATGATGATGATGATGATGATG 3318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2253 CCATGCCCTCTCTACAGTGTGTGAACACGAGCTGCTCTGTACAAAGATGATGATGATG 2312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3319 CCCATGACAGACCAAGTCTTCAA 3343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2313 CTGAGAGCTGAGAGCTTCCCAA 2337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 2  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-F1s  
US-08-232-463-14

Query Match 0.9%, Score 57.2, DB 1, Length 7218;  
Best Local Similarity 1.3%; Pred. No. 2.7e-05;  
Matches 5; Conservative 229; Mismatches 142; Indels 0; Gaps 0;

```
OY 4907 TTCCCTTTCCTTACACCTGTCATCTGTCTCTATTTCTGGAATACGTTCTCCCT 4966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1060 TTGCATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 4967 GACCTGCTGATACAGGAGGCTCTCTCAAGTACTGTTGAACGTCGGCTTTTGTG 5026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1120 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 5027 TAGCTCCACCCACTGTAGAGGCTAGCTGCTTTGAAGAACTCCCATTTGCAAC 5086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1180 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 5087 GGACCCGCGCCGCGCAGACTGTTCACAAAGTTCCCGCCCAACCCAGACATCA 5146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1240 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 5147 GCTGTAGCTCCCTGCTGAGGAGTGTATGTTCCAGAGTGGGGCTCAGAGCC 5206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1300 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 5207 CTCTCCAGAACTTTAGTGGCCCTCTACCTGACTCTGATCTTTTACAG 5266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1360 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 5267 TAGCTTCTCCCTCG 5282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 YYYYYYYYYYYYYYG 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3  
US-08-838-219B-20  
Sequence 20, Application US/08838219B  
Patent No. 5877012  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Natini M

```

: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: A No. 5877012el Class of Proteins for the
: TITLE OF INVENTION: Control of Plant Pests
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/838,219B
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,483
: FILING DATE: 06-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Gary M.
: REGISTRATION NUMBER: 40,403
: REFERENCE/DOCKET NUMBER: CGC 1925
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8582
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2241 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "synthetic DNA encoding
: : HYPOTHEICAL: NO
: : US-08-838-219B-20

Query Match 0.7%; Score 46; DB 2; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 846 ATCCGGGAGTGGTGAGTGAAGAGCTCTCTGACAGTGGAGCGCCAGCGCAGAGCTGAC 905
DB 160 AACCAAGCAGCTGCTGAACAGATAGCGGCAAGCTGGAGCGCGTGAAGCGACCTGAC 219
OY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAGCTGTAGAGAGAGCATTGAGAACGCC 965
DB 220 GACCTGATCGGCCAGCGCAACTGACACGACGAGCTGAGCAAGAGATCTTAAGATGCC 279
OY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAAGCCAGACCGACATTTCTACAAGAG 1025
DB 280 AACGAGCAAGCAGCTGCTGAAGAGCTGGAACAACAGCTGAGAGCGCATCAACACCATG 339
OY 1026 CT 1027
DB 340 CT 341
```

```

RESULT 4
US-09-233-336A-20
: Sequence 20, Application US/09233336A
: Patent No. 6107279
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Kozziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: A No. 6107279el Class of Proteins for the
: TITLE OF INVENTION: Control of Plant Pests
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/233,336A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/838,219
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Gary M.
: REGISTRATION NUMBER: 40,403
: REFERENCE/DOCKET NUMBER: CGC 1925
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8582
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2241 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "synthetic DNA encoding
: : HYPOTHEICAL: NO
: : US-09-233-336A-20

Query Match 0.7%; Score 46; DB 3; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 846 ATCCGGGAGTGGTGAGTGAAGAGCTCTCTGACAGTGGAGCGCCAGCGCAGAGCTGAC 905
DB 160 AACCAAGCAGCTGCTGAACAGATAGCGGCAAGCTGGAGCGCGTGAAGCGACCTGAC 219
OY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAGCTGTAGAGAGAGCATTGAGAACGCC 965
DB 220 GACCTGATCGGCCAGCGCAACTGACACGACGAGCTGAGCAAGAGATCTTAAGATGCC 279
OY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAAGCCAGACCGACATTTCTACAAGAG 1025
DB 280 AACGAGCAAGCAGCTGCTGAAGAGCTGGAACAACAGCTGAGAGCGCATCAACACCATG 339
OY 1026 CT 1027
DB 340 CT 341
```

Db 160 AACACAGCTGCTGAACGACATCAGCGGACCTGGACGCGCTGAACGCGACCTGAAAC 219  
QY 906 AGGAGATGCGCCAGCTGATCTGAAGATCGAAGCCTGTAAAGAGACATTTGAAGACGCC 965  
Db 220 GACCTGATGCGCCAGGCAACCTGAAACAGGAGCTGAGAGAGATCTTAAGATCGCC 279  
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAGCAGACCGACGATTCCTCAAGAG 1025  
Db 280 AACGAGCAACACAGGTGCTGTAACGACGTGACACAGCTGACCGCATCAACACCATG 339  
QY 1026 CT 1027  
Db 340 CT 341

RESULT 5  
US-09-233-752A-20  
Sequence 20, Application US/09233752A  
Patent No. 6137033  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233.752A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2241 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA encoding  
DESCRIPTION: VIP3A(c)"  
HYPOTHETICAL: NO  
US-09-233-752A-20

Query Match 0.7%; Score 46; DB 3; Length 2241;  
Best Local Similarity 53.3%; Pred. No. 0.017;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGATGAAGAGCTCTGACAGCTGAGGCCAAGGCCAAGCTGAAC 905  
Db 160 AACACAGCTGCTGTAACGACATCAGCGGACAGCTGAGCGGCGTGAACGCGACCTGAAC 219  
QY 906 AGGAGATGCGCCAGCTGATCTGAAGATCGAAGCCTGTAAAGAGACATTTGAAGACGCC 965  
Db 220 GACCTGATGCGCCAGGCAACCTGAAACAGGAGCTGAGAGAGATCTTAAGATCGCC 279  
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAGCAGACCGACGATTCCTCAAGAGAG 1025  
Db 280 AACGAGCAACACAGGTGCTGTAACGACGTGACACAGCTGACCGCATCAACACCATG 339  
QY 1026 CT 1027  
Db 340 CT 341

RESULT 6  
US-09-402-036-20  
Sequence 20, Application US/09402036  
Patent No. 6291156  
GENERAL INFORMATION:  
APPLICANT: Yu, Cao-Guo  
APPLICANT: Estruch, Juan J.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Kozziel, Michael  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284C  
CURRENT APPLICATION NUMBER: US/09/402.036  
FILING DATE: 2000-02-08  
PRIOR APPLICATION NUMBER: PCT/EP98/01952  
FILING DATE: 1998-04-02  
PRIOR APPLICATION NUMBER: 08/838,219  
FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,263  
FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,265  
FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/463,483  
FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/314,594  
FILING DATE: 1994-09-09  
PRIOR APPLICATION NUMBER: 08/218,018  
FILING DATE: 1994-03-24  
PRIOR APPLICATION NUMBER: 08/037,057  
FILING DATE: 1993-03-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 20  
LENGTH: 2241  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA  
OTHER INFORMATION: encoding VIP3A(c)  
US-09-402-036-20

Query Match 0.7%; Score 46; DB 4; Length 2241;  
Best Local Similarity 53.3%; Pred. No. 0.017;



FILED DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2370 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA encoding  
DESCRIPTION: VIP3A(b)"  
HYPOTHETICAL: NO  
US-09-233-336A-19

Query Match 0.7%; Score 46; DB 3; Length 2370;  
Best Local Similarity 53.3%; Pred. No. 0.017;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGTGATGAGAGCTCTGAGCTGAGAGCCGCAAGAGCTGAC 905  
DB 160 AACGAGGAGCTGTGATGAGAGCTCTGAGCTGAGAGCCGCAAGAGCTGAC 219  
QY 906 AGCGATGCGCAAGCTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCC 965  
DB 220 GACCTGATGCGCAAGCTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCC 279  
QY 966 AAGCAGGAGCTGTGAGAGCTCTGAGAGCTCTGAGAGCTGAGAGCTGAGAG 1025  
DB 280 AACGAGGAGCTGTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCTGAGAG 339  
QY 1026 CT 1027  
DB 340 CT 341

RESULT 9  
US-09-233-752A-19  
Sequence 19, Application US/09233752A  
Patent No. 6137033  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the  
TITLE OF INVENTION: Control of Plant Pests  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233,752A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2370 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA encoding  
DESCRIPTION: VIP3A(b)"  
HYPOTHETICAL: NO  
US-09-233-752A-19

Query Match 0.7%; Score 46; DB 3; Length 2370;  
Best Local Similarity 53.3%; Pred. No. 0.017;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGTGATGAGAGCTCTGAGCTGAGAGCCGCAAGAGCTGAC 905  
DB 160 AACGAGGAGCTGTGATGAGAGCTCTGAGCTGAGAGCCGCAAGAGCTGAC 219  
QY 906 AGCGATGCGCAAGCTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCC 965  
DB 220 GACCTGATGCGCAAGCTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCC 279  
QY 966 AAGCAGGAGCTGTGAGAGCTCTGAGAGCTCTGAGAGCTGAGAGCTGAGAG 1025  
DB 280 AACGAGGAGCTGTGATGAGAGCTCTGAGAGCTGAGAGCTGAGAGCTGAGAG 339  
QY 1026 CT 1027  
DB 340 CT 341

RESULT 10  
US-09-402-036-19  
Sequence 19, Application US/09402036  
Patent No. 6291156  
GENERAL INFORMATION:  
APPLICANT: Estruch, Juan J.  
APPLICANT: Yu, Cao-Quo  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Kozziel, Michael  
APPLICANT: Nye, Gordon  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284C  
CURRENT APPLICATION NUMBER: US/09/402,036  
CURRENT FILING DATE: 2000-02-08

```
; PRIOR APPLICATION NUMBER: PCT/EP98/01952
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 08/838,219
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,263
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,265
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/463,483
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/314,594
; PRIOR FILING DATE: 1994-09-09
; PRIOR APPLICATION NUMBER: 08/218,018
; PRIOR FILING DATE: 1994-03-24
; PRIOR APPLICATION NUMBER: 08/037,057
; PRIOR FILING DATE: 1993-03-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; OTHER INFORMATION: encoding VIP3a(b)
; US-09-402-036-19

Query Match          0.7%; Score 46; DB 4; Length 2370;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTCGGTGAAGAGCTCTGACGTGGAGGCCCAAGCGCAAGCTGAAC 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 aaccagagctgctgtaacgcacacgcagcagctgagcgtgtaacgcagcctgaac 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 906 AGCGAGATCGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGAGAACGCC 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 gacctgctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 966 AAGCAGGACCTGCTCCAGCAAGATGTGATGAGCCAGCCGCTTCTCAAGAGAG 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 aacgagagagaaacagtgctgtaacgagcgtgtaacaaagcgtgtaacaaacacacatg 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1026 CT 1027
    ||
    340 ct 341

RESULT 11
; US-08-471-033-30
; Sequence 30, Application us/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Marsha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note="maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3a(a)"
; US-08-471-033-30

Query Match          0.7%; Score 46; DB 1; Length 2403;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTCGGTGAAGAGCTCTGACGTGGAGGCCCAAGCGCAAGCTGAAC 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 176 AACCAGAGCTGCTGTAACGATCAGCGGCAAGCTGAGCGGCTGAAGCGCAGCTGAAC 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 906 AGCGAGATCGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGAGAACGCC 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 236 GACCTGATCGCCAGGAGCAACCTGAACACCGAGCTGAGCAAGAGATCCTTAAGATCGCC 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 966 AAGCAGGACCTGCTCCAGCTCAAGATGTGATGAGCCAGCCAGCAAGCTTCTCAAGAGAG 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 296 AAGCAGGAGCAAGCAAGTGTGTAAGAGAGTGAACAAACAAGCTGACGCAATCAACACCATG 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1026 CT 1027
    ||
    356 CT 357

RESULT 12
; US-08-471-044-30
; Sequence 30, Application us/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Marsha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
```

APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note="maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-471-044-30

Query Match 0.7%; Score 46; DB 2; Length 2403;  
Best Local Similarity 53.3%; Pred. No. 0.017;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGAAGAGCTCTGACCTGAGGAGCCCAAGGCGCAAGAGCTGAAC 905  
DB 176 AACACAGCAGCTGCTGAACATCAGCGGCAACCTGACGCGCTGAAGCGCACCTGAAC 235  
QY 906 AGGAGATGCGCAAGCTGAATCTGAAGATCGAAGCTGTAAAGAAAGCATTTGAGAAAGCC 965  
DB 236 GACCTATGCGCCAGGAGCAACCTGAGACACCGAGCTGAGCAAGAGATCTTAAGATCGCC 295  
QY 966 AAGCAGAGCTGCTCCAGCTCAAGATGTATCAGCCAGCAGCAGCATTCCTACAGGAG 1025  
DB 296 AACGAGCAACACAGTGTGTAACACAGCTGAGCAACAGCTGAGCCATCAACACCAAG 355

QY 1026 CT 1027  
DB 356 CT 357

RESULT 13  
US-08-463-483A-30  
Sequence 30, Application US/08463483A  
Patent No. 5849870  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozief, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,483A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8615  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note="maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-463-483A-30

Query Match 0.7%; Score 46; DB 2; Length 2403;  
Best Local Similarity 53.3%; Pred. No. 0.017;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;



1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

```

? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? HYPOTHETICAL: NO
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 11..2389
? OTHER INFORMATION: /note= "maize optimized DNA
US-08-471-046A-30 sequence encoding VIP9(a)"

Query Match 0.7%; Score 46; DB 2; Length 2403;
Best Local Similarity 53.3%; Pred. NO. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGAGTGAAGAGCTCCTGACGCTGGAGGCCAAGCCAGAGCTGAAC 905
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 176 AACCGAGAGCTGCTGAACGACATCAGCGGCAAGCTGGAGCGGCTGAAGCGGCGAGCTGAAC 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 906 AGCGAGATCGCCAACTGTAATCTGAAAGATTCGAAGCCTGTAAAGAAAGCATTTAAGAACCC 965
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 236 GACCTGATCGCCCAAGGCAACTGTAACACCGGAGCTGAGCAAGAAATCTTAAAGATCGCC 295
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 966 AAGCAGAGACTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGAGCAATTCCTACAAGAG 1025
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 AAGCAGAGCAACCAAGGTGCTGAAGACGCTGAACACAAACAAAGCTGGAGCGCATCAACACATG 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1026 CT 1027
   ||
DB 356 CT 357

RESULT 15
US-08-470-566B-30
? Sequence 30, Application US/08470566B
? Patent No. 5872212
? GENERAL INFORMATION:
? APPLICANT: Warren, Gregory M
? APPLICANT: Koziel, Michael G
? APPLICANT: Mullins, Martha A
? APPLICANT: Nye, Gordon J
? APPLICANT: Carr, Brian
? APPLICANT: Desai, Nalini M
? APPLICANT: Kostichka, N. Kristy
? APPLICANT: Duck, Nicholas B
? APPLICANT: Estruch, Juan J
? TITLE OF INVENTION: No. 5872212a1 Pesticidal Proteins and Strains
? NUMBER OF SEQUENCES: 52
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 5872212a1s Corporation
? STREET: 3054 Cornwallis Road
? CITY: Research Triangle Park
? STATE: NC
? COUNTRY: USA
? ZIP: 27709
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08470.566B
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/463,483
? FILING DATE: 05-JUN-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/314,594
? FILING DATE: 09-SEP-1994

```

```

1  PRIOR APPLICATION DATA:
2  APPLICATION NUMBER: US 08/218,018
3  FILING DATE: 23-MAR-1994
4
5  PRIOR APPLICATION DATA:
6  APPLICATION NUMBER: US 08/037,057
7  FILING DATE: 25-MAR-1993
8  ATTORNEY/AGENT INFORMATION:
9
10 NAME: Meigs, J. Timothy
11 REGISTRATION NUMBER: 38,241
12 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SQLV44
13 TELECOMMUNICATION INFORMATION:
14
15 TELEPHONE: 919-541-8587
16 TELEFAX: 919-541-8689
17
18 INFORMATION FOR SEQ ID NO: 30:
19
20 SEQUENCE CHARACTERISTICS:
21     LENGTH: 2403 base pairs
22     TYPE: nucleic acid
23     STRANDEDNESS: single
24     TOPOLOGY: linear
25
26 MOLECULE TYPE: other nucleic acid
27
28 DESCRIPTION: /desc = "Synthetic DNA"
29
30 HYPOTHETICAL: NO
31
32 FEATURE:
33
34     NAME/KEY: misc.feature
35     LOCATION: 11..2389
36     OTHER INFORMATION: /note="maize optimized DNA
37     OTHER INFORMATION: sequence encoding VIP3a(a)"
38
39 US-08-470-566B-30

```

		0.7%;	Score 46;	DB 2;	Length 2403;
Query Match		Best Local Similarity	53.3%;	Pred. No. 0.017;	
Matches	97;	Conservative	0;	Mismatches	85; Indels 0; Gaps 0;
QY	846	ATCCGGAGTGCCTGAGTGAAGAAGCTCCTCGACGTGAGGCCAAGCGCCAAAGCTGAAC	905		
Db	176	AACCCAGCTGCTGTGAACGACATCAAGCGGCAAGCTGAGCGGCTGAACCGCACCTGAAC	235		
QY	906	AGCGATGCGCCAAAGCTCAATCTGAAGATCCGAAGCGCTTAAGABAGCATTTGAGAACGCC	965		
Db	236	GACCTGATCGCCAGGGGCAACCTTGAAACCGAGCTGACCAAGGAATCTTTAAGATCGCC	295		
QY	966	AAGCAGGACTGCTTCACAGCTCAAGAAATTTCATACGCCAGACGAGCATTTCTACAAAGAG	1025		
Db	296	AACGAGCAAGAACAGGTCTCTAGACGACTGAACAACAAAGCTGGAGCGCATCAACACCATG	355		
QY	1026	CT 1027			
Db	356	CT 357			

```
Search completed: September 8, 2002, 22:22:30
Job time: 12741 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 18:03:19 ; Search time 590.89 Seconds  
(without alignments)  
17933.619 Million cell updates/sec

Title: US-09-809-920-3  
Perfect score: 6172  
Sequence: 1 GCGGGCTCCTGAGCTGAA.....ATMAATGAGACTTAACTC 6172

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0  
Searched: 1736436 seags, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6172	100.0	6172	21	AAA12734
2	5833	94.5	5854	20	AAV72378
3	2459.6	39.9	3479	21	AAA12733
4	2347.4	38.0	3198	22	AAE26844
5	2212.4	35.8	2253	21	AAC59000
6	882.2	14.3	1599	22	AAE26843
7	596.4	9.7	598	22	AAE68305
8	545.6	8.8	3595	23	AB126683
9	545.6	8.8	6105	23	AB126682

10	310.4	5.0	316	14	AA061373	Human brain Express
11	267.8	4.3	756	20	AAV89868	EST clone IE352.
12	207.2	3.4	353	20	AAV89056	EST clone C31. Ho
13	200.6	3.3	3008	21	AAE21849	Human breast and o
14	200.6	3.3	3175	20	AAE08224	Human Ext-2 CDNA.
15	199	3.2	3003	19	AAV19375	Hereditary multipl
16	169.4	2.7	3183	20	AAE08223	Human Ext-1 CDNA.
17	166.2	2.7	2698	23	AB109865	Drosophila melanog
18	166.2	2.7	3642	20	AAE08225	Drosophila melanog
19	163	2.6	3817	23	AB161229	Drosophila melanog
20	160.2	2.6	2965	24	AB199788	Mouse ischemic co
21	159	2.6	176	16	AAE24258	Human gene signatu
22	137.4	2.2	4733	23	AB110422	Drosophila melanog
23	137.4	2.2	4820	23	AB109864	Drosophila melanog
24	90.2	1.5	60747	23	AB16128	Drosophila melanog
25	68.4	1.1	617	23	AAE94481	DNA encoding novel
26	52.4	0.8	6976	22	AAE52949	Human polynucleoti
27	52.4	0.8	7569	22	AAE51965	Human polynucleoti
28	51	0.8	51	22	AAE17426	Human silent SNP c
29	51	0.8	51	22	AAE17428	Human silent SNP c
30	50.2	0.8	23449	21	AAE35393	Maize amylose-exte
31	50	0.8	50	22	AAE17430	Human silent SNP c
32	49.8	0.8	905	21	AAE77845	Human cancer assoc
33	49.4	0.8	51	22	AAE177427	Human silent SNP c
34	49.4	0.8	51	22	AAE177429	Human silent SNP c
35	49.2	0.8	42299	22	AAE68932	Human immune/haema
36	48	0.8	8323	24	AB132058	Human immune syste
37	47.8	0.8	5987	24	AB133563	Human immune syste
38	47.2	0.8	10732	21	AAA10594	Gene encoding a su
39	47	0.8	8136	24	AB132554	Human immune syste
40	47	0.8	99960	21	AAE50905	Human TBC-1 patila
41	46.8	0.8	7516	24	AB132060	Human immune syste
42	46.8	0.8	17967	24	AB133014	Human immune syste
43	46.6	0.8	18301	22	AAE198944	Human excretory re
44	46.6	0.8	18301	22	AAE163294	Human kidney relat
45	46.4	0.8	965	22	AAE51884	Human polynucleoti

## ALIGNMENTS

RESULT 1  
ID AAA12734 standard; CDNA: 6172 BP.

AC AAA12734  
XX 25-JUN-2000 (first entry)

DE CDNA encoding human TREX protein.

KM Tumour necrosis factor receptor-associated Factor; TRAF;  
KW TRF-protein-interacting hereditary multiple exostos protein; TREX;

KW signal modulator; tumour necrosis factor receptor;

KW CD40 mediated signal transduction; TRAF protein; cancer;

KW hereditary multiple exostos; autoimmune disease; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 594..3353 /\*tag-a

FT Product "tumour necrosis factor receptor-associated Factor (TRAF) protein-interacting hereditary

FT multiple exostos (TREX) protein"

PN 06-APR-2000.

XX 17-SEP-1999; 99WO-US21654.

PK 17-SEP-1998; 98US-0156191.

*Applied*



QY	1681	TCGAAAGAGAAATGAGAGCGGACCCCTCCCGCGACTACGATGACCGGATTCATTGGCACCC	1740
DB	1681	TCGAAAGAGAAATGAGAGCGGACCCCTCCCGCGACTACGATGACCGGATTCATTGGCACCC	1740
QY	1741	TGGAAGGGGTGACAGACAGCAACCTGGATTCAGTCTCTGGTAATTCCACTGCAAAAACC	1800
DB	1741	TGGAAGGGGTGACAGACAGCAACCTGGATTCAGTCTCTGGTAATTCCACTGCAAAAACC	1800
QY	1801	AGCCCAAAACCCAGGCTCCGACACTGAGTGGGACACTGTTGGTAAGAGCGGAGACCGCTTGG	1860
DB	1801	AGCCCAAAACCCAGGCTCCGACACTGAGTGGGACACTGTTGGTAAGAGCGGAGACCGCTTGG	1860
QY	1861	AATTGCTGAAGCTCTCCACTTCGCCCCATCATTAACCCCGGAGACCCCTGTTGGTTA	1920
DB	1861	AATTGCTGAAGCTCTCCACTTCGCCCCATCATTAACCCCGGAGACCCCTGTTGGTTA	1920
QY	1981	TGCTGGGGGAGACAGCTCAGCTTCCTCCATCAGAGACATGCTGCAGTGGAGAGAGGCGGCC	2040
DB	1981	TGCTGGGGGAGACAGCTCAGCTTCCTCCATCAGAGACATGCTGCAGTGGAGAGAGGCGGCC	2040
QY	2041	TGGTGTGGCCAAAGCCCTGCTTACCGAGGTTCAATTCTGCTCAGAAAGCCTTCCTCCATA	2100
DB	2041	TGGTGTGGCCAAAGCCCTGCTTACCGAGGTTCAATTCTGCTCAGAAAGCCTTCCTCCATA	2100
QY	2101	GTGACCTCTCGGCTATATAGAGCGGCGCAAGGCGGCTTCTCTGGGAGACTTACTTCTCCACTG	2160
DB	2101	GTGACCTCTCGGCTATATAGAGCGGCGCAAGGCGGCTTCTCTGGGAGACTTACTTCTCCACTG	2160
QY	2161	CTGACAGTATTTTTAATACCGTCTGGCTATGATTAGACTGCACTGCATCCAGATCCAGCCG	2220
DB	2161	CTGACAGTATTTTTAATACCGTCTGGCTATGATTAGACTGCACTGCATCCAGATCCAGCCG	2220
QY	2221	CTCCCATCCCGGGAAGAGCGGCGACGTGAGATCCCCACCGTTCAAGGCAAGGCGGTGGAA	2280
DB	2221	CTCCCATCCCGGGAAGAGCGGCGACGTGAGATCCCCACCGTTCAAGGCAAGGCGGTGGAA	2280
QY	2281	CTGACCCCAATGAGCTGCAACAAGGGGAGCTGGACCTGGGGCGACATGGAGACAGAGCCG	2340
DB	2281	CTGACCCCAATGAGCTGCAACAAGGGGAGCTGGACCTGGGGCGACATGGAGACAGAGCCG	2340
QY	2341	CCTAAGCCCTCACCAGATACCTCCGCAATTTCATCTGACTGTCACTGACTTTTACCGCA	2400
DB	2341	CCTAAGCCCTCACCAGATACCTCCGCAATTTCATCTGACTGTCACTGACTTTTACCGCA	2400
QY	2401	GCTGGAAGTGTGCTCCAGGGGCTTCCTCANTTTTCCCCACACACTCCCTTACCCCTGTGT	2460
DB	2401	GCTGGAAGTGTGCTCCAGGGGCTTCCTCANTTTTCCCCACACACTCCCTTACCCCTGTGT	2460
QY	2461	TGCCCTCAGAGGCCAATTTTGTGGGCTCAGAGGACTGGCTTGTGGGCTATTGGTGTGTAG	2520
DB	2461	TGCCCTCAGAGGCCAATTTTGTGGGCTCAGAGGACTGGCTTGTGGGCTATTGGTGTGTAG	2520
QY	2521	CTGGGGGTTCTGGCAAGGAATTTTCAGGCAAGCGTTTGAGGCAATGTTCCCGAGAGCAGT	2580
DB	2521	CTGGGGGTTCTGGCAAGGAATTTTCAGGCAAGCGTTTGAGGCAATGTTCCCGAGAGCAGT	2580
QY	2581	TCAGAGGTGTGATGTTTGCATTAAGAGCGGAGGAGTAAGTGTATTGAACCTTTAAGAGAGC	2640
DB	2581	TCAGAGGTGTGATGTTTGCATTAAGAGCGGAGGAGTAAGTGTATTGAACCTTTAAGAGAGC	2640
QY	2641	TGAATGGGCTCTCTTACTCTCAACAAAGGTGCTGTGTGTAATTTCCCAAGCTGCAT	2700
DB	2641	TGAATGGGCTCTCTTACTCTCAACAAAGGTGCTGTGTGTAATTTCCCAAGCTGCAT	2700
QY	2701	CAGAGGACCTTCTGTGGCTGTGACATTGGGCTTCCATCATGGTGTCCGTACTGAGAAGA	2760
DB	2701	CAGAGGACCTTCTGTGGCTGTGACATTGGGCTTCCATCATGGTGTCCGTACTGAGAAGA	2760

QY	2761	ACAGTTTGAACACACCATT	TTTCTTACCTGGAA	TAATTAAGACAGGCAT	CTGTGTCA	2820
Db	2761	acagtttgaacacaccgat	tcttacccttggaa	taataatbgaacagagccat	ctctgtcca	2820
QY	2821	TTGATGACGATGCTCACCTTC	CGGCATGAGAAATCA	TGTTTGAGTTCGGGG	GTGGAGAG	2880
Db	2821	ttgatgacgaatgctcaccttc	cggcatagaaatcatgtttg	agtttcgggggtggagag		2880
QY	2881	AAGTCGGGAGCCGATG	TGGGCTTCCTTGCCG	GTTCACACCCATG	GGACATCCGCAATC	2940
Db	2881	aagtcgggagccgatg	tgggcttcccttggccgtt	acacccatgggacatntccgc	aatc	2940
QY	2941	AGTCCTGGCTCTACACT	GTCACTGTGAGCTGT	CCATGCTGTCA	GAGTGCTG	3000
Db	2941	agtcctggctctaaactct	caactactcctgtgagct	gcatacgtgtcgtacag	tgctg	3000
QY	3001	CGTTCTTTACCAAGTAT	TATGCGCTACCTGAT	TTCCTTATGATG	TGATGATG	3060
Db	3001	cgttctttaccaaagtattat	gctgacctgacattctta	tgatgattgacacagccat	ccgg	3060
QY	3061	ACATGTGTGATGAT	TACATCACTGTGAGAC	ATTGCCATGAACTT	GTGTCTCCACA	3120
Db	3061	acatgtgtgatatgata	ctacactgtgagacatgtc	gaatcgtacatctctgtc	ctccaca	3120
QY	3121	TCATCTCGGAAGCCCC	CCCATATCAAGTGAC	CTCACGGTGGACAT	TTCCGATGCCAGAT	3180
Db	3121	tcatctcggaagcccccat	aatgataagcttcaacgt	gtgacatctcgaatgc	atgccccagatgc	3180
QY	3181	CTCAGGCCCTTCTCAT	GATGATGACCTCCAC	ACTTCCACAGGGGACA	AAATGACATCAACT	3240
Db	3181	ctcaggcccttctcatg	atgataagcttccacag	gtgacaggtgacaaatgc	atctctc	3240
QY	3241	TCGTGAAGGTGTAC	GCGCTACATGCCCT	CTCTGTACACGCAG	ATTACAGGATTC	3300
Db	3241	tcgtgaaggtgtacgct	acatgccctctctgtac	acgcagattacaggttctg	tc	3300
QY	3301	TCCTTCAGACACGCT	CTCCCCATGACAA	GCACAGTGTTC	CAAGTTCATAGGGG	3360
Db	3301	tcttcaagaacagctctc	cccatgacaagaagct	gtctcaagtctcatg	gtggcagc	3360
QY	3361	GCACGGTCTGGGGA	AGATGATGACAGGAG	GAAGATGCGTCC	CAAGTTCCTAGCA	3420
Db	3361	gcacggtctvgggaag	atgatagacaggaagaa	tgatcccaaggtctcc	aaagca	3420
QY	3421	TTGCGAGACCTTTGG	GCACATCTGCTGG	TGGTGCCCAAGGCT	CTGCTGGAAGGG	3480
Db	3421	ttgcgagaccttggca	catctgtcgttgggtg	ggccagagctctgtc	gtgaaaggag	3480
QY	3481	CAGAGCAGTGGAA	GAAACCGCTGCTT	ATCTTAAGTAC	GCACACTGGGCT	3540
Db	3481	cagagagatbgaag	aaacagctgtcctat	ctbgaatcagcca	acatgtggcc	3540
QY	3541	CCCTGGGGGAGTCC	CGCGGGGTTCCCA	CACAGGAGACT	GCATGATGTTAC	3600
Db	3541	ccctgggggaggtcc	cggggttccccaacag	gtgacatgataagct	ctaacacag	3600
QY	3601	ACTGTGGGACTTGC	AGACTTCACACAG	CTGTTCTCTAC	AGCCAGACAGCTG	3660
Db	3601	actgtgggactctg	caagatctaacacag	ctgtctagcccaag	acagctgtctg	3660
QY	3661	GTTTTTACATCAAT	TAACAACATTTAT	GATTTATTTAA	AAAGAAATTTCA	3720
Db	3661	gttttacaatcaat	taacaacatlbata	cttlaaaaaagaa	agtlttcagat	3720
QY	3721	ATTCAAGCCTTAT	TATATATATGTG	TGTATATATAA	ATACATGCAAC	3780
Db	3721	attcaagccttata	tataataatggtgtat	ataataacatgca	acacatgtcata	3780
QY	3781	TATATATTTGGCT	GGGAGAGTGTAG	TTTTCCTTTTAA	GAGGAGGAGCCG	3840
Db	3781	tataataatttggct	gggaggtgtgagt	tttctcttcaagg	aggaacgcgcag	3840
QY	3841	TCCTTTGTCTG	ATCTGTGCGGAG	ATGGGTCTG	GCCTTGCTGTG	3900

Db	3841	tcctttgtctcgtatctccgagcgaatgaggtcccgccctcgtgtaactcgtctcctta	3900
OY	3901	AAGATCATCTCCCATCTCCACAGGCGCATCTGTGTGACGACACAGAAAGATGAAC	3960
Db	3901	aaatcatctccatctcccccagcgacatctgtgtgtagaacaacagaaagatgtaact	3960
OY	3961	TGGCCCTCTTTCGGGGCCCTGGACAAGGTCTCTCTTACCCCTTCTGTGTCAGCA	4020
Db	3961	tggccctctcgtcgccctcgaagagtcctctcctaacccttctcgtgcagctcagaa	4020
OY	4021	CCTGTAACTCATCTTCTTCTCCAGTGAATCCCTGGGGGCGCCGACCCCTGTGGCTGT	4080
Db	4021	ccgttaactcaacatctctctctccacgtgaaaccgcggagacgcctgagccctgtcgt	4080
OY	4081	TCACTTCTCTGCTCTGGGGCCACGCAATTTTGAAGATTTATCTTGAAGCCAGCTTGC	4140
Db	4081	tcaactctcgtcgtcgaggccagcgttcttgagttatcttagccagcttcgc	4140
OY	4141	TCCGTACTTATCCCTGCTCTCCCATTTTCTCTCTTGTGAGAGAGAAATGAGGAAGCAAG	4200
Db	4141	tccgtactctctccgtcctccctcccatctctctctgttttgagagagaaatgagaaagcaag	4200
OY	4201	AGTGAGAAAATATGAGGGCTGAAGACCCACCTCCCAATGCGCTTCTTACCCGCTCT	4260
Db	4201	agtgagaaagaatagaggcctggaagacgcaactcccgaatgacctctctcctgcctc	4260
OY	4261	CTGTGAAACACACGTGCTGTGGGCTCAGGCGTTTCTGAAGTGCCTTCTTGATTTGG	4320
Db	4261	ctgttgaaacaacacgctcgttgaggccctcagcgcttcttgaaagtccctctctgatatg	4320
OY	4321	ACGAGGATATGAGAGGCTGACATCTCTGTGCTGTAATGTGTCAGGTACAGTACGCTCC	4380
Db	4321	acggagatcatagacagcgtgacaacctcgtcgttgctctgaatgtgttgcaagctacgctcc	4380
OY	4381	TCTCCCTAGTGTAGAGCAACCATGTCTCTTCTGAGGAACCCACCGCGCTGGCGGGAGT	4440
Db	4381	tctccctagtgtagagcaagcaggttccctctgaggaaccacccggcctgagccgggaagt	4440
OY	4441	TTTACAGCAAGGCGCTGCCCTTGGGATTAATTCCTTGGTGAATTCACCTTCCCGCGCT	4500
Db	4441	tttacagcaagcgctcgtccttgagtaatctctcgttgaaatcaactctcccccgcct	4500
OY	4501	CTGTGAGGCCCCATCTCGTGTATCTGTCGTTTTTGGACCCCTAATGTACAGTTGGCT	4560
Db	4501	cgctcggagcccacaccccggtgtactcgtgtgttcttgacccctaagtctcagcttgcct	4560
OY	4561	GTAGAGCTCCCGAGGTTTGATGTGTGTAACAATGAGAGGCTGTGATTTGCTGTGA	4620
Db	4561	gttagagctcccagagttctgtgtatgtgcataagaaatggaggtcgtgtgttctgcgtgta	4620
OY	4621	AAGTCATCCACGCTTGGAAATCTAACGGGCAATTCACAAACCGAGTTACACTTTCAC	4680
Db	4621	aaagtcacatccagcctctgtaatctaacgacattcaacaacccagttaccacttccact	4680
OY	4681	CCCTGTTTGGAGTTCGTTCCCTGGGCGTAAACTAAATAACTAATTTTGGGTACAG	4740
Db	4681	ccctgcttaagatctcgttccctcggccgaaactgaaataagcttaatttctgggtcag	4740
OY	4741	GTCGCAATGAGGGAACCTAGAGAGGTGTGAGGCAATTTGTACAGGATTTACCCATGAC	4800
Db	4741	gttgcaagtacgggaacctcagagaggtgtgagtgacattgtcaaggaatttagccatgac	4800
OY	4801	GTCGTTCTTGAACCTTACTTTCTGGAAATGAGATGTGACCTTGGAATTTCTACAACTG	4860
Db	4801	gtgtctctgaacctactctctctgtaagtgtgacatctcggaaagttctctagcaactg	4860
OY	4861	AACAAACACTCAGGTTTGTCTCGGTATGACACATGCTCTTAAACCAAGTTCCGTTCCCTA	4920
Db	4861	aacaaaactcaggttctcgtgtcagtcacaatgccttaagccagttccgtctcccta	4920
OY	4921	GACCTGGCATCTGTGTTTCTGATTTCTTGAATATACGTTTCTCTGTACCTGCTGTACC	4980
Db	4921	gacctggcatctgtgttctgatttcttgaatatcgtttctctgtacctgctgtacc	4980

[illegible]

RESULT 2  
 AAV72378 standard; cDNA: 5854 BP.  
 AAV72378.  
 29-JUL-1999. (first entry)  
 Human EXOSTOSIN-4 CDNA.  
 Exostosin-4; human; cytosolic; immunosuppressive; anti-HIV; Ext-2; treatment; cancer; tumour; genetic disorder; AIDS; diagnostic; screening; acquired immune deficiency syndrome; ss.  
 Homo sapiens.  
 Key Location/Qualifiers  
 CDS 263..3022  
 /\*tag= a  
 /product= "Exostosin-4"  
 WO925822-A1  
 27-MAY-1999.  
 14-NOV-1997; 97WO-CN00126.  
 14-NOV-1997; 97WO-CN00126.  
 (U9HU-) UNITV-HUMAN MEDICAL.  
 Deng H, Fan C, Liu C, Ruan Q, Xia J, Xu L.  
 WPI: 1999-347470/29.  
 P-PSDB: AAY08481.  
 New Exostosin-4 gene useful in the treatment of cancer, tumors, genetic disorders and AIDS  
 Claim 2; Page 22-24; 33pp; English.  
 This invention describes a novel human Exostosin-4 polypeptide which has cytosolic, immunosuppressive and anti-HIV activity. Exostosin-4 has homology with Ext-2 and is therefore expected to have similar biological activities. Exostosin-4 and its modulators can be used for treatment of cancer, tumours, genetic disorders and AIDS (Acquired Immune Deficiency Syndrome), and in diagnostic/screening assays for such conditions.  
 Sequence 5854 BP; 1204 A; 1611 C; 1611 G; 1428 T; 0 other:  
 Query Match 94.5%; Score 5833; DB 20; Length 5854;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 5836; Conservative 0; Mismatches 5; Indels 0; Gaps 0.  
 332 GGCACCTTTATCGTTTGAATTAAGACATGTTCTTGGTCAACAGCAGCAACTT 391  
 1 ggcacccattatcgcgtttgtaagaagattaaagagcagatgcttttgcgtcaacagcagaact 60  
 392 AAATCTGCTGGAATAGGGTCAAGACCACTTTCAGCTGAGTGAAGAAATGAATGTT 451  
 61 aaaaatcgtcgtgaatagggcagagacatttcagctgagctgaggaatgaatgatt 120

QY	452	CATTATTATTTGGTCCCTTGTGCTGGGAGACACTATACCTCTTGTGGAACGTGTGCTGTA	511
Db	121	catttatttvgtgcccttgcgcggagacacatacctctctggaaacgtgtccagtgaa	180
QY	512	ACACAGATCGTTTGTGTGGATTATCAACCCATGTTTATGGCAGTAGTACCGCAGCTATCTG	571
Db	181	acagagatcglttctgtgtaabtagcaaccatggtctatvgcgagtgaccgcagtgactctg	240
QY	572	GGGGGACGCTGCAGAGAGACTCATGACAGGGCTATPACCATGTGCGGAAATGGGGCGCGGG	631
Db	241	gggggacgagctgcagagagactcaatgaacgctatcacatgctgcgaaatvggggagcgcg	300
QY	632	GAACGAGAGTCAAGACCTGCGATCTGTGGCGTGTGCCAACCCGATCCGCTACCTGGCTCAG	691
Db	301	gaacgagagtgtaagacgctgcgtctgtgcgtgttccaaacgcatctgcgtctaaagtggtcag	360
QY	692	CTTACACCTCTTGTCTACTGCTGTGCTCTTCTCCGCTCATCGGCCACTATPACTCTACAC	751
Db	361	cttaacgctcttgtcatctctgtcttctcttccgcctcatcgcacatacttaacctaacac	420
QY	752	TCTGATGAGCTGATGAGGACAGGCAAGCGGATTTTGTGCTCCCGGTGGGGAGAGACT	811
Db	421	tctgtagtgagcgtatgtaggcaggaagcgagatcttctgtctcccggtgtgggaaagagct	480
QY	812	GTCGAGAGTGAAGCACTGCTGATCTGTGTCCGCATCCGGAGTGGGTAGTGAAGACT	871
Db	481	gtcgagagtgaagacagtgctgtgtagctgtgcgcgcatccggagtgctgtatgtgaagagct	540
QY	872	CTGTGACGTGGAGGCCAAGGCGCAAGCGTGMAACCGAGATCGGCAGAGCTGAATCTGA	931
Db	541	ctgtgcagctgtagagccaaagcgccaaagagctgaaacagcgagatctgcgaagctgtaa	600
QY	932	GATGGAAGCCGTGAAGAAAGACATTAGAAGACGCAAGCAGAGACTGTCTACACTCAAA	991
Db	601	gatacgaagcctgtaaagaaagacatitagaagcgcaagacgaagcgtctccagcttaagaa	660
QY	992	TGTGATCAGCCACAGACGAGCATTTCTTACAAGAGCTCATGCCCCAGAACCCCAAGCT	1051
Db	661	tgatcatagccaaagccgaagatcttccaaagagctcatgtgcgaagaaacagcccaagct	720
QY	1052	GTCCTGTCCATCCGACTGTCTCCAGAAAGACGATGGCGGCTGCCCTCCCGGCAAGC	1111
Db	721	gtccctgtcccatctcgagctgtctcccaagaaagcgtctgcggcctctcccccgcgaagc	780
QY	1112	CACTCGGGGCGCCGCGCTATACAAACGTGTTGAATTTCTGTCGCCCTCACCTCTGG	1171
Db	781	cactcggggtgcgcgggtacaacaacgtcttgaattctctgttgcctctcaactctcgg	840
QY	1172	CTTCCCGGTCTACGTCTATGATGACAGTGAACGATTTGTCTTTGGAGCGTACCTGATCCCT	1231
Db	841	cttcccggtctacgtctctatgacagtgacaggtgtgtcttctgtgcagctaacctgtacctc	900
QY	1232	GGTCAACGAGGCTTTTAGGCGACAGCAGAGCTTAACGTTTATGTTACGAATAATGCA	1291
Db	901	ggtcaacgaggtcttctaagcgacgacgacgactaagcttctgttaacgaaatgycaga	960
QY	1292	CATGCGCCGCTTATACGTGATPACAGTGAAGGAGAGTGAAGAGCCGCTGTCTCTCGGC	1351
Db	961	catgcgcgtccttaagtgtaactagtgtggagagatgacgaagagccgggtgtcgtcggcc	1020
QY	1352	TGCTGAGCTGGAGAACGATGTGTATTCCTGTCACACTGGCGGAGGAGATGACACACAA	1411
Db	1021	tgctgagcttgagagaaagatgtatctccctgcacacgttggcggaagatgagacaaacca	1080
QY	1412	TGTGATCATCATCTGTGTACAGTAAAGTCAGATACACGAACCTTCTCTATACGTACAGTAC	1471
Db	1081	tgatcatcatcatctgtcgaatgaatgaatgaatacaagaacctctctataaagctaaagc	1140
QY	1472	TGGCGGAGCAATGGTGGCCCAAGTCAACGCTTCTACATGTGTCAGTACAGACCTGGCTTGA	1531
Db	1141	tggcggagcagcaatgtgtgcgcagctcctctctacacgtgtccagtgaaagccgggttttga	1200
QY	1532	CTTGTGTATCACCGCTGTGCTCATGCAATGTTGTGAGCCCAACTTCATGGAATCCAC	1591



Db 1201 cttggtcgtatcaaccgctggtccatgctcgtgacccaacttcagtgaatccacc 1260  
QY 1592 ACAGGGGCGCGTGAACGGAAATATCTTTCACCTTCAGGGCGAGAAAGTTGACTCT 1651  
Db 1261 acagggtccggtgaaacggaatatctctcacctccaggcgagagaatltgactctc 1320  
QY 1652 GAAGGTAGCCCTCAGAGAGCGCGCTCTCTGAAAGAGAAATGAGGGGACCCCTCCGC 1711  
Db 1321 gaggctacgctccagagagcccgctcctctgaaagaaatgagggccctccgc 1380  
QY 1712 CGACTACGATGACCGGATATTCACCTGAAAGCGGTGCAGAGACGAAGCTGATCA 1771  
Db 1381 cgactacgatagcgagatcatltgccacctgaagcggtgcagagacaagctgataca 1440  
QY 1772 GGCCCGGTGGGAATTACCTGGAAAAACAGCCAAACCCAGCTCCGATGATGGCG 1831  
Db 1441 ggtccggtggaattacattcagaaacacagcccaaccacgctgccaactgagtggc 1500  
QY 1832 ACTGTGTGAGAGCGGAGACCGCTGTGGAATTGCTGAAGCTCTCCACCTGCGCTCAT 1891  
Db 1501 actgtgtgagagcgagagacgctggaattgtctgaagctctccacctgcgctcat 1560  
QY 1892 CATTAACCCCGGAGACCTCGCTGTGTTATTTCTCTGGGTGCAACAGCGCTCTTCA 1951  
Db 1561 cattaacccccgggagaccctgctgttattccctctggtgtgacaacgctctcga 1620  
QY 1952 AGCCCTGGAAATGCGGTGCGCTGCGTGTGTGTGGGGAGCGATCCAGCTTCCTACCA 2011  
Db 1621 agccctgaaatcggtgcgtccgctggtgtgtgtggtgagcgagctccctcacca 1680  
QY 2012 GGACATGCTGAGTGAAGCAAGGCGCCCTGTGTGTGCAAGACCTCGTTACCGAGGT 2071  
Db 1681 ggaatgctgagtgagaaagcgcgccctgtgtgtgtgcaaaagctctgttaccgaagt 1740  
QY 2072 TCATTTCCTGCTCAAAAGCTTCCGATATGATGACTCTGCTAGTATGAGCGGCAAGCG 2131  
Db 1741 tcaattctctgctcaaaagctctctcgaatgactcctctgtgtatgagcgcgcaagcg 1800  
QY 2132 CTTCCTGTGGGAGACTTACTTCTCCACTGCGTGAAGTATTTTATATACCGTCTGCTAT 2191  
Db 1801 cttctctcgtgagacttactctctcaactgctgaagtaattttaaaccgtgtcgtcat 1860  
QY 2192 GATTAGACTGCGATCGATGCCAGCCGCTCCATCCCGGAGAGGCGGAGCTGAGAT 2251  
Db 1861 gattagactgctgcatccagatcccaagcgctcccaatccggaagagcgagctgagat 1920  
QY 2252 CCCCCACCGTTCAAGCAAGGCGGCTGGAACTGACCCCAACATGGCTGACAAAGGGACCT 2311  
Db 1921 cccccacggttcagagcaagcggtgtgaaactgaaacccaacatggtgtgcaacggtgacct 1980  
QY 2312 GGACCTGGGCGCATGAGAGGAGCGCGCTAGCGCTCACCCAGATATACCTCCGCAATTT 2371  
Db 1981 ggaactggggtgcagatgagagagcgccctacagctccacacagatacctccgcaatt 2040  
QY 2372 CACTCTGACTGTCACTATTTTACCCAGCTGGAACCTGCTCCAGGGCTTTTCACTCT 2431  
Db 2041 caactctgactgactacttacttaccgagctggaactgtgtctccagggctcttccatct 2100  
QY 2432 TTTTCCCCACACTCCCTTTTGAACCTGTGTGCTCTCAGAGGCAAAATTTTGGGCTCAG 2491  
Db 2101 ttcccccaactcctcttgacctgtgtgtccctcagagccaattcttggctcag 2160  
QY 2492 GACTGCTTTCGGGCTATTTGTGTGTGAGCTGGGGGTTCGGCAAGAAATTTTCAGCAGC 2551  
Db 2161 gactgcttccgacctactgtgtgtgtgagctggtgtgtgtgtgtgtgtgtgtgtgtgt 2220  
QY 2552 GCTTGAAGCAATTTTCCCGAGAGAGCTCAGGTTGATGTTGCTTATGAGCGGGA 2611  
Db 2221 gcttgagagaaatgttcccgagagagcaatcaggtgtgtgtgtgtgtgtgtgtgtgtgt 2280  
QY 2612 GGAAGTGTATGAACTCTTTAGAGAGGCTGAATGGCTCTCTTACTGAAAGAGTGTGT 2671  
|||||

Db 2281 ggaagtgtctatgaactctttagagaggtcgaatggtccctcccttaccatgacaagtgct 2340  
QY 2672 GGTGTGTGGAATTTCTCCAAAGCTGCCATGAGAGACCTTCTGTGGCGTGAATGGCGT 2731  
Db 2341 ggtgtgtggaattctcccaagctgtcccaagctcagacagagacctctgtgtgtgtgtgtgt 2400  
QY 2732 TCCATCATGTGTGCTCCGTACTGAGAAAGAACATTTGAACCAACGATTTTACCTCGAA 2791  
Db 2401 ccccatcatgtgtgtcgttactgtagaagaagcttgcgaacaacgcttcttaccctgga 2460  
QY 2792 TGAATTTGAGACAGAGGCGCATCTGTGCAATTTGATGAGTATGCTACACTCCGATAGCA 2851  
Db 2461 tgaatttgaacagagagccatccctgtccatltgatagactgtccactccgcacatgacga 2520  
QY 2852 AATCATGTTTGGGTTCCTGGGTGTGAGAGAACTCGGAGCCCATGCTGGGCTCTCCCTGG 2911  
Db 2521 aatcatgt 2580  
QY 2912 CCGTTACACGATGGAGCATCCCATGATGATGCTGTGCTTACAACTCCAACTACTCTGT 2971  
Db 2581 ccgttacacagatgagacatcccatcagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2640  
QY 2972 TGAGCTGTGATGCTGACAGGTGCTGCTCTTCTTTCAGATATATGCTTACTCTGTA 3031  
Db 2641 tgaagcttccatgagtgagacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2700  
QY 3032 TTCTTATGTATGCCCAAGGCCATCCGGGACATGCTGTGATGAATACATCAACTGTGAGA 3091  
Db 2701 ttcttatgtatgt 2760  
QY 3092 CATTCGATGAATCTCTTGTGTCTCCCAATCACTGTGAAAGCCCCCATCAAGTGAAGCTC 3151  
Db 2761 catgtccatgaactctctgt 2820  
QY 3152 ACGGTGACATTCCTCCATCCAGATGCCAGATGCCCTGAGGCGCTGATGATGATGCTCCACTT 3211  
Db 2821 aaggtgacatctcgaatgtcccaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880  
QY 3212 CCACGAGGCGCAAGATGATCAACTTCTGTGTGTAAGAGGCTACATGACCCCTCT 3271  
Db 2881 ccaagagtgagcaaaagtgtcatcaacttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2940  
QY 3272 GTACAGGAGTTCAGGAGGTGATTTCTGTCTCTTCAAGACACAGCTGCCCCATGACAAAGC 3331  
Db 2941 gtacagagatgtcaggt 3000  
QY 3332 CAAGTGCTTCAAGTTCACTAGGGGAGCGACAGCTGTGGGAGAGATGAGAGAGAGG 3391  
Db 3001 caagtgtctcaagtcatctcaggt 3060  
QY 3392 AGGAAGATGCTCCCAAGGTTCCTAGGCAATTCAGAGACCTTGGGACATGCTGTGTGGG 3451  
Db 3061 aggaagatgt 3120  
QY 3452 TGGCCGAGGCTCTGCTGGAAGGGGAGCAGAGAGAGTGAAGAGAAACCGCTGCTTTA 3511  
Db 3121 tggccgagagctctgt 3180  
QY 3512 TCTTGAATGAGCAGCACACTGGGCTGTGAGCCCTGGGGAGTCCCGGGGTTCGCCACAC 3571  
Db 3181 tcttgaagtgaagccacacgt 3240  
QY 3572 AGGGCACTGACTATGACTTACACTGAGACTGTGGGACTGTGAGAGTCACTCAACAC 3631  
Db 3241 agggcaactgtactgtacttaccagt 3300  
QY 3632 GTTCGTACGCCGAGAGAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3691  
Db 3301 gtctgtacgcccgagagacaggt 3360  
QY 3692 ATTTAAAGAGAGAGTTTGAATTTGCCATTCGAAGGCTTATTTATATATGTGTGTGT 3751  
Db 3361 atttaaaagagaaagcttcaagatttgcacatcaaggttatatatatatgtgtgtgtgtgtgt 3420

OY 3752 ATATAAATACATGCACACTTGGCATATATATATTTTGGCTGGGGAGCTGTGAGTTT 3811  
|||||  
Db 3421 aataaataacatacgaacaactctgacataataatattcttgccgcyggggagctgagcttc 3480  
OY 3812 TGCCCTTCTAAGGAGGAGGACCGCGCAGGCTCCTTTGTTCTGTATTTCTGGCGGAATGGGT 3871  
|||||  
Db 3481 tgccttcttaaggaggagacgcgcagagctcccttgctctgtaattcttgcyggaagtgtt 3540  
OY 3872 CCTGGCCTTGTGTACATGCGCTTATCCTTTAAAGATCATCTCCCATCTCCCGCAGCGCATC 3931  
|||||  
Db 3541 ccctgaccttgctgaaccgctgctataccttaagaatacctcccatcctcccccagcgccatc 3600  
OY 3932 TGTGTGAGCAACCAAGAAAGGAGTGAATTTGGCCCTTGGCGGCGCTGGACAGGCTCTCT 3991  
|||||  
Db 3601 tctgtgcagcaacacagaaaggagatgaactctgcccctcttgcyggccctgagaagctcctc 3660  
OY 3992 TCCCTTACCCCTTCTGTGTGGCCAGTCAAGCAACTGTATCTCATATTTCTCTCCAGTGAATC 4051  
|||||  
Db 3661 tcccttaacccttctctgtctgcaagcaaacctgtaactcaatctctctccagctgaatc 3720  
OY 4052 CCTGGAGAGCGCTACCCCTGTGGCTGTGAGCTTCTCTGCTGGCGGCGCAGGATTTT 4111  
|||||  
Db 3721 ccctgggagcgccctgaaccctggctggctgtcaagcttccctgctctgcyggccagagcttc 3780  
OY 4112 TGAGGATTTATCTTTAAGCCAGGCTTGCCCTCGTACTTATCCCTGCTCTCCATTTCTCT 4171  
|||||  
Db 3781 tgaagatttaattcttaagccagagctgcccctcgtactatctccctgctcccatcttcc 3840  
OY 4172 CTGTTTGGAGAGAAATGAGAGCAAAAGAGTGAAGAAAGATAGGGCTGAAGACGCCAC 4231  
|||||  
Db 3841 ctcttcttgaaggagaaatgaaggaaacaaagagctgaagaaagaaatgaaggcctgaagcgcac 3900  
OY 4232 TCCAGATAGCGCTCTTTGTATCTCTCTCTCTCTGTTGAAACACACCTGCTGTGGGCTCAG 4291  
|||||  
Db 3901 tcccagaatgaccttctctatctctctctctctctctctctgtaaaacaacgctgctgggctccag 3960  
OY 4292 CGTTTCTGAACTGCTCTTTCTTGGATTTGGACAGAGATCAGACGCTGACATCTGCTGT 4351  
|||||  
Db 3961 cgtctctgaagctgct 4020  
OY 4352 GGTGTGAAGTGGTTTGGAGGAGGCACTCCTCCCTAGTGAAGAGCAAGCCAGTGCCTT 4411  
|||||  
Db 4021 ggtctgaaggctgtctctgaagctcaagctctctccctctctctctctctctctctctctct 4080  
OY 4412 CGAGGAACCCACCCGCGTGGCGGGGAAGTTTACAGCAAGGCGCTGCTTGGGATATT 4471  
|||||  
Db 4081 cgaggaaaccacaaccgagctggccgggaagttttacagcaaggcgccctctgggataatc 4140  
OY 4472 CCTTGGTGAATTTACCTTCCCGCGCTCTGTCTGGAGCCCATCTGTGTATCTGTG 4531  
|||||  
Db 4141 ccttggctgaattcaacctctcccccgcctctgctctgagcccaatccctgttctatctgtg 4200  
OY 4532 GTTTTGGAGCCCTAATGTACGTTGGCTTGAAGATCCTCCGAGGTTTGGATGTGCTAG 4591  
|||||  
Db 4201 gttcttggaaccctcaagctgaagctgtgctgtgaagaccccccgaagctctgtatctgtcag 4260  
OY 4592 AACAAATGGAGGCTGTGATTTGTGTGTAACTCATCATCAGCCTTGGAAATCTAAGGGC 4651  
|||||  
Db 4261 aacaaatggaggctgtgattctgtctgttgaagctcaaatccagccttgaaatctaaagggc 4320  
OY 4652 ATTACAAACCCGAGTTACCATTTCCACATCCCTGCTTAGATTTCTGTTCCCTGGGCTGAA 4711  
|||||  
Db 4321 attacaacaaccgaattcaaatctcaatctccctgctctgaagattctgttcccttgaggctaa 4380  
OY 4712 ACTGAATTAAGCTAATTTTGGGTCAGGTCAGGTCAGTGAAGGAACTGAGAGGCTGTAG 4771  
|||||  
Db 4381 actgaataaagctaatattcttggtctatgtgtgaagctagggaaacctaggggggtctgtag 4440  
OY 4772 TGGCAATTTGTACGGGATTTAGCCCATGACGTGTTCTTGAACCTTACTTCTTGAAGTGG 4831  
|||||  
Db 4441 tggcaattgtcaggagattctagcccatgagctgttctctgaaaccttaattcttgaaaggg 4500

OY 4832 AGTTGACTGTGGAAGTTTCTAGCAACTGAAACAAACCTAGAGTTTCTCTGTGTCATGCA 4891  
|||||  
Db 4501 agttgactctggaagttctcttagcaactgaaacaaagctcaagttcttccctgtgcaatga 4560  
OY 4892 CATGCTTTAAGCCAGTTCGCGCTTCCCTAGACCTTGGCATCTGCTGCTTCTTCTTCTGG 4951  
|||||  
Db 4561 catgctttaaagccaagctctctctccctgaacctctggaacctctgtgtctcaattctcttg 4620  
OY 4952 AATGACTTCTCTGTGACACGTCGCTGTACCAAGTGGTCTCTTCAATAGTGTTTTGAAG 5011  
|||||  
Db 4621 aataagctctccctctgacctgcgcctgtataccaagctgtgtccctctcaagtactgtttgaag 4680  
OY 5012 CTGGGCTCTTTTGTGTAGCTCCACCCACCTGTAGGGCTAGCTGCGCTTAAAGGAACTCT 5071  
|||||  
Db 4681 ctggacctcttctgttagctctccaccaccactgtgaagctgaagctcgtgtcttaaggaaactc 4740  
OY 5072 CCCCATTTGGAAACCGGACCGCGCGCGGCGCAGGACGTGTTCCAAAGGTCCCGCGCC 5131  
|||||  
Db 4741 ccccatctgcaaaccggaaccgc 4800  
OY 5132 CCAACCCGACATCAGCCTGTAGCTCCCTGCTGAGGACAGTGTGTTATGTTCCACAG 5191  
|||||  
Db 4801 ccaaccccgacatcagcctctgtagctccctctgtctgaaggcaatgtgtatgttcccgagcag 4860  
OY 5192 TGGGCTCAGACGCCCTTCTCTAGAACTTCTTACTTCCCTTACCTGACTCTGACTTG 5251  
|||||  
Db 4861 tgggctcagaacgcgcctctcccaaatcttctgaatctgtcgtctccagctccctgactgt 4920  
OY 5252 TATTCCTTTTAGCAGTGTAGCGCTTCTCCCTGGGGAGCAAAAGATGTGTGTGGGCT 5311  
|||||  
Db 4921 tatctctttagcaagctctctctccctccctcgcgggagccaagaaatgtgtgtgagcgtc 4980  
OY 5312 ATATTGTGCGTGTATTTCAATCTGCTGTTCTTTTAAATGTGAAGAACTCATATCTGACTTC 5371  
|||||  
Db 4981 atattgtgcgtcatttcaattcaattctgttctttaaigtgggaactcaaatcaatcagacttc 5040  
OY 5372 AGTGGACTCGGTAGACCGCGGGCGCGTCTGTGTGGTGGGACCCCTTTAAGGGACTAGT 5431  
|||||  
Db 5041 agtggactcgtgtgagccggggccgtctgtgtgtgtgagacccctcttaaggggagctagct 5100  
OY 5432 GAGCTGGGCGCGTGTGTGTGTGAGGCAAGGCGCTTCCCTTAAATGTGAAGCAAGTGTGTC 5491  
|||||  
Db 5101 gagctggggcgctctgt 5160  
OY 5492 GGGCCCGAATGTCATGTTGTGATCTAAGAAAGGCTGAGTGTGTCACCAAAACATGC 5551  
|||||  
Db 5161 gggcccgaaatgtcaactgtgtatctaaggaggctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5220  
OY 5552 CGCAGGAGGCGTGTGTGTGTGCGCGGTCTTCCAAACAAGAACAGCCCTCTGTGAAAG 5611  
|||||  
Db 5221 cgcaggaggagctgt 5280  
OY 5612 GAACACTGTGTTGAAGAACGTGACAGACAGGCTGTAGGGGACGCGCTCTCAGCGAAGAG 5671  
|||||  
Db 5281 gaacaactgtgttgaagaaatcgtgaagaacagctcttgaagggtgcaagcctccctcaaggagaag 5340  
OY 5672 CAGCAAGTGGCCACAGTGTACTGTGTGAGGTCTTACACACGGGAAAGCGCGCAGC 5731  
|||||  
Db 5341 cagaagaagtgtgccaagctgtcaactgtcgaagtgtctctcaacaaggaaagccgcgcagcc 5400  
OY 5732 TGTGACTGCGTGTGAATGGGAAAGCGGCGCACAGACCCGCGGTCTCTTGGTGTGCTGT 5791  
|||||  
Db 5401 tgtgaactcgtcttgaagatgaggaaacggcgcaacagaaacccgggtctctctgtgcgtctgtc 5460  
OY 5792 GGGCGCGCCCTGGCCACCTTGTCTGTGCTGTGCAAGGTGTGCAAGAGCGGCTGTCTGGG 5851  
|||||  
Db 5461 gggcgccgccccgtgcacactgtctctgtctcgaagggtgcaaggagcgtctgtctctcggg 5520  
OY 5852 TGGCGGCTTGTGCTCGGCTCGGTTTGGGCTGTCTTACCAATTAACACCGTCCAGGGCTGTGA 5911  
|||||  
Db 5521 tggcggcttgtctgtctcgcgt 5580  
OY 5912 GGGCACTGTGAGCGGCTGCGTCCCTGGGCAAGTGTCTCTCCGTGTGACTGTGTGCTCAAGGC 5971  
|||||

Db 5581 ggcactgtgagcgtgcctcctggcagtgctctccgtgtgagctgctcagcc 5640  
 QY 5972 AGGGCTCACAGCTGGGGTCTCTCCGGAAGATGGATCTTCTGGAGCTGCGCGGA 6031  
 Db 5641 agggctcacagctgggtctctccggaagatggatctctcggagctgcgcgga 5700  
 QY 6032 CAGAGTGGGAGAGCTTCTAGTTTGGGGGACCTTGATATCCATGCCACGTCATCCA 6091  
 Db 5701 cagagtgaggagctcctagtgttggtgggaaagcttgatccatgcacgcacatca 5760  
 QY 6092 CCCACCCCTTTTGTCTCAGCAGCAGCAATGCTTACATTGATTTTGTAAAAATATA 6151  
 Db 5761 cccaccccttctgcacagcagacaatgctctacattgatttctgtaaaaaataa 5820  
 QY 6152 AATAATGAGACTTAACTC 6172  
 Db 5821 aataatgagacttaactc 5841

## RESULT 3

ID AAA12733 standard; cDNA; 3479 BP.  
 XX AAA12733;  
 XX

DT 25-JUL-2000 (first entry)

DE cDNA encoding mouse TRFX.

XX Tumour necrosis factor receptor-associated Factor: TRAF;

KW TRF-protein-interacting hereditary multiple extoses protein; TRFX;

KW signal modulator; tumour necrosis factor receptor;

KW CD40 mediated signal transduction; TRAF protein; cancer;

XX hereditary multiple extosis; autoimmune disease; ss.

XX Mus sp.

OS location/Qualifiers

XX Key 458..3214

XX CDS /tag= a

XX /product= "tumour necrosis factor receptor-associated

XX Factor (TRAF) protein-interacting hereditary

XX multiple extoses (TRFX) protein"

XX W0200018959-A1.

XX 06-APR-2000.

XX 17-SEP-1999; 99MO-US21654.

XX 17-SEP-1998; 98US-0156191.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI: 2000-293180/25.

XX P-PSDB; AAY84668.

XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated

XX Factor (TRAF) protein-interacting hereditary multiple extoses (TRFX)

XX protein, useful in the diagnosing cancer -

XX Claim 20; Fig 7A-B; 161pp; English.

CC is cancer, a hereditary multiple extosis or an autoimmune disease. The  
 CC cancer is colon cancer, gastric cancer, human head and neck squamous  
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,  
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,  
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,  
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.  
 XX  
 SQ Sequence 3479 BP; 762 A; 959 C; 951 G; 806 T; 1 other;

Query Match 39.9%; Score 2459.6; DB 21; Length 3479;  
 Best Local Similarity 84.2%; Pred. No. 0;  
 Matches 2902; Conservative 0; Mismatches 505; Indels 40; Gaps 10;

QY 163 CTGTTCTAATCTTACTTATTAAGTGAATAGACATCATATTTGTGGTCAGCA 222  
 Db 53 cgtttttaccagccgctgcatcact-tyaataagaagcatgcatatggtcgcgaca 111  
 QY 223 AAACCAAGAACAGAGCTATGCAATTTGAAAAAGCTGTGTGATTCAGAGGTGTTTTC 282  
 Db 112 aagccaagygacaaaagctatgcccgttaaatgctcctgagtcagggctcttcc 171  
 QY 283 CTGGGTTTCATCAGTACGTAACCTCCCTTCATCTCAGCAAGATGTGACCTTTTA 342  
 Db 172 cggcttctaagacacatgcatctccttccatccatcagcaatggtacccctt 231  
 QY 343 TCGTTTGATTAAGATTAAAGACATGTCTTGTCTCAACAGCAGCACTTAAATCTGCTG 402  
 Db 232 ctacttgatgag-----aaagctgatacttcagattgctc 268  
 QY 403 GAATAGGCTCAGACACCAATTCAGCTGACGTGAGAAATGTAATTTATTTT 462  
 Db 269 gactaaggtagaacccgtaalcgctg---tgaagaaatgaaattccatttactg 324  
 QY 463 GTGCTGTCTGGGGAGAGACACTAATCTCTGGAAGCTGCAATGGAAGAGATGCT 522  
 Db 325 gtgcctgtgcaaggaagacacatgctcctccagaactgtgtgtgaa-agaagttgc 383  
 QY 523 TTTGTGAATAGCAACCCATGTTATGGAGTAGACCCGAGTGATGCTG---GGGGGAG 579  
 Db 384 gtttgcagaagacatgactatggtatagcgagcgaatccgaagtatacgaagtggcaaga 443  
 QY 580 GCTGACAGAGACTATGACAGGCTATACCATGCTGCGGAATGGGGCGGGGAACGAG 639  
 Db 444 ggcacagcgaactcatgacagcgtatacactgtgcaatggtggaggtgggaagcgtg 503  
 QY 640 GTCAACCTGCTGCTGCGGCTGTCACACCGCATCCGCTACCTGCTGCTGCTGCTG 699  
 Db 504 gtcaacctgtatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 563  
 QY 700 TCTTTGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 759  
 Db 564 tgtcatcatcctctgcttcttccctcatatgctatctatctacacacatcttgagc 623  
 QY 760 AGGCTGATGAGGACGAGCGGATTTTGTGCTCCCGGTGGGAGAGAGCTGTGCGAG 819  
 Db 624 aggcagaagagctgagcaagcgtatctcgccctggcgtgagcgtgagcgtgagcgtg 683  
 QY 820 TGAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 879  
 Db 684 taaagctgtccttgaatctctgctgctgctgctgctgctgctgctgctgctgctgctg 743  
 QY 880 TGAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 939  
 Db 744 tctgaagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 803  
 QY 940 CTGTGTAAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999  
 Db 804 cctgtaagaagagatagagatagcagaagcagcagcagcagcagcagcagcagcagcag 863  
 QY 1000 GCGAGACCGGAGATTTCTTAAGAGAGCTGATGGCCCAAGAACAGAGAGAGAGAG 1059  
 Db 864 gccagaagagcactctactaagaagcgtgagtcgccaagacagcccaactgtccctgc 923

[illegible]

QY	2140	GGAGACCTTACTTCTCCACTGCTGACAGATTTTAAATACCTGCTGGTATGATTAGA	2199
Db	2001	ggaagaccactactctccacccgcagacagatattttaataccgtgtcgtgcgaatga	2060
QY	2200	CTCGCATCCAGATTCACAGCCGCTCCCATCCGGAGAGAGCGGCACCTGATGCCCAAC	2259
Db	2061	ctcgaaattcagaatcccaagctgtctcccaatcccggaagaggttagcggtctgaaatcccccaatc	2120
QY	2260	GTTCAGGCAAGCGGCTGGAACTGACCCCAACATGGCTGTACAAAGGGGACCTGTGACCTGG	2319
Db	2121	gttcaagcaaaagcagctggaactgaccccccaaatcgtctgaacaatgggagcctgagaccgtg	2180
QY	2320	GGCCAGTGGAGACGAGAGCCGCCCTACGCTCTACACCAAGATACCTCCGGCAATTTTACTCTGA	2379
Db	2181	ggccggttagagacaagaaccaacctatgtcctcaactaaatattcctcgcgaatttcaactctga	2240
QY	2380	CTGTACTACTCTTTAACCCACACTGGAAATGTGCTCCACAGGGGCTTTCATCTTTTCCCCC	2439
Db	2241	ctgtcaagaactgttaaccgttgagctggaactctgcgcccggaagcaggttccaatcttlttcccc	2300
QY	2440	ACACTCCCTTTTGACCCTGTGTGGCCCTCAGAGGCGCAAAATTTTGGGCTTCAGGAGACTGGCT	2499
Db	2301	acaacaccttgaatcctgtgtgtccctctgagggccaaattctctgggtctcaaggactgagt	2360
QY	2500	TTTCGGCTATTGGTGTGGAGCTGGGGGTTCTTGCGCAAGAAATTTACAGCAGCGCTTGGAG	2559
Db	2361	tttcggtcgaatcgtgtgctggtggtgtctgggcaagagatgttccaagcagcgtctcgag	2420
QY	2560	GGAATGTTCCCGAGAGCGTTACAGGTTGGTATGTGACTTATGAGCGGGAGGAAGTGC	2619
Db	2421	ggaatgttccccgagagcggttcaaggttcaaaagtttgaatgttatttgccttcccccttccccctt	2480
QY	2620	TTATGAACCTTTTAAAGAGGCTGAATGGGCTCCCTTACCTGTGAACAAGTCTGTGGTGTGT	2679
Db	2481	tcatgaactccctctgagagagatcacaagcgtccctcccaacttgaaacaaggtgagtgtgtgtgt	2540
QY	2680	GGAATTTCTCCAAAGCTGCCATATGAGAGCACTTCTGTGGCTACATTTGGCGTTCCCATCA	2739
Db	2541	ggaactctcccaagcgtccctctgagagacatttctgtgcagaacatttgtgtctcccatca	2600
QY	2740	TGCTGGTCCGTACTGAGAAACAGTTTAACACCGATTCTTACCCTGGATGAAATTTG	2799
Db	2601	tgtgtgtccgttaactgtgagaaacagtttgaacaaatcgtgtctctgcgcgtgaatctgaatctgt	2660
QY	2800	AGACAGAGCCATCCCTGTTCATTTGATGAGAGTCTACCTCCGCCATGACGAAATCATGT	2859
Db	2661	agacagagagcactactgttcatcagaagatgagtctcaactctcgccatctgaatgaatcatgtt	2720
QY	2860	TTTGGGTTCCGGGTGTGGAGAGAGCTCGGGACCGCATGCTGGGGCTTCCCTGGCGCTTACC	2919
Db	2721	tttgggtttcgggtgtgtgagagagaacgtgtacgtcatctgtgtgtgtctccctcgtgcgttacc	2780
QY	2920	ACGCATGGGACATCCCCCATCAAGTCTGTGGCTCTACAACCTCAACATCACTTCCTGTGAGCTGT	2979
Db	2781	atgcgtgtgagacatccgcgacacagctccctgtctctcaaatctcaactctcgtgtgagctgt	2840
QY	2980	CCATGGTGTGACAGGTGTGCTGCTTTTCAACAAGATATATACCTCAACCTGTATTTCTTATG	3039
Db	2841	ccatggtgtcgtacgagcgtcgtctctctcttcaaaagtattatagtcctcaactgttcttaagt	2900
QY	3040	TGATGCCCGCAGGCATCCGGGACATGGTGGAATATCATCAACTGTGAGCAATTGCCA	3099
Db	2901	tgatgccccagagcatalccggtgagatgtgtgagaggtatacaaatctgttgaggtatctcgca	2960
QY	3100	TGAACCTTCTTGTCTCCCATCTCGGAAGCCGCCCATCAAGTGAACCTTCAAGGTGGA	3159
Db	2961	tgaactctctgtctctcccaatcaaacggaaccccccaatcaaggltgaatcaaaagtgtga	3020
QY	3160	CATTCCGATGCCAGAGATCCCTCCAGGGCCCTGTTCATATGATATGATCCCACTTCCACAGGAC	3219
Db	3021	cattctcgaatgcacaaagggtgcctctcagggccctgttcccaatgaatgaactctcaatcttcaagagc	3080
QY	3220	GGCACAAGTGCATCAACTTCTTCTGTGAAGGTGTACGGGTACATGGCCCTCTCTGTACAGCC	3279

```

Db      3081 ggcacaaagtatcaacttttgcaggtgtaagcgtatatacctctcttcttgaacac 3140
QY      3280 AGTTCAAGGGGATTTCTGTGCTTTTAAGACACGCTGCCCATGACAAACCAAGTGT 3339
Db      3141 agtctcgggtgagctctctcttcaagaccgctgcctccatgacaagaacaaagtgt 3200
QY      3340 TCNAATTCAATCTGAGGCGACGCGTGGGGAGAGAGATAGACAGAGGAGGAAGAT 3399
Db      3201 tcaagtctctagagcctcttcagttctgagagacaatgagacg-aggagggggagtc 3259
QY      3400 GCGTCCCAAGTTCTTAGGCAATGACGACCTTGGGACATCTGCTGTGGTGGCCAG 3459
Db      3260 accctcaaggtccccaaggtctcgaaggtccttgaggacatctg-tcggcaggggcacg 3318
QY      3460 AAGCTCTGCTGGAAGGGGAGGAGGAGGAGTGAAGGAAACCGCTGCTT-ATCTTGA 3518
Db      3319 acccttgcgtgagagagagagagagagagagagagagagagagagagagagagagag 3378
QY      3519 GTACGACACACTGGGCGTGGAGCGGCGGAGTCCCGGGGTTCCCAACAGAGGAC 3578
Db      3379 gtcaagcacaactggcgctggagatcctctgacagagactcagncgt--ctgcacagggcac 3436
QY      3579 TGACTGATAGCTTACACTGAGACTGT 3605
Db      3437 tgactgatacgcaacactgagactgt 3463

```

## RESULT 4

AAF26844 ID AAF26844 standard; cDNA; 3198 BP.

AAF26844;

05-APR-2001 (first entry)

Rat Reg-binding protein encoding cDNA SEQ ID NO:3.

Rat; Reg-binding protein; Reg receptor; Reg; DNA synthesis; apoptosis; cell proliferation; diabetes; antidiabetic; ss.

Rattus norvegicus.

Key Location/Qualifiers  
 CDS 5..2764  
 /tag= a  
 /product= "Reg-binding protein"

W0200077192-A1.

21-DEC-2000.

09-JUN-2000; 2000MO-JP03764.

10-JUN-1999; 99UP-0164488.

(OKAM/) OKAMOTO H.

Okamoto H;

WPI; 2001-061871/07.

P-SDB; AAB51341.

Protein binding to pancreatic Reg protein and its encoding DNA, useful for screening candidate antidiabetic drugs -

Claim 1; Page 64-74; 11pp; Japanese.

The present invention describes DNA which encodes: (a) a protein of rat origin binding to Reg protein; or (b) a Reg-binding protein derived from the natural Reg sequence by addition, deletion and/or substitution of one or more amino acid residues. The present sequence encodes a rat Reg-binding protein from the present invention. Reg-binding proteins

CC have antidiabetic activity. Drug compositions comprising Reg-binding CC proteins and their genes are useful for the treatment and/or prevention CC of diabetes. In addition, they are useful in developing new treatments CC for diabetes.  
 xx  
 SQ Sequence 3198 BP; 723 A; 882 C; 850 G; 743 T; 0 other;

Query Match 38.0%; Score 2347.4; DB 22; Length 3198;  
 Best Local Similarity 85.3%; Pred. No. 0;  
 Matches 2721; Conservative 0; Mismatches 416; Indels 54; Gaps 7;

```

QY      590 ACTCATGACAGCGTATACCATGCTCGGAGTATGGGGCGGGGGAAGAGAGTACAGCTG 649
Db      1 actcatgacagcgtatataccatgttgcgagatggggaggggagagagagagagagagag 60
QY      650 CATGGCGCGCTGGTCAACCGCATCCGCTCAACGTCAGTGGCTCAGCTTCCAGCTTTGTAT 709
Db      61 tatggcgcgctgggtcaacccgcatccgctgagacctgagagttcaagctgttcatcat 120
QY      710 CCTGCTCTCTCTCCGCTCATCGCCACATTAATACCTCAACACTGTGATGAGAGCTGATGA 769
Db      121 cctgctctctctcccgctcatcgccacatatactcaaccactcggatgagagacatga 180
QY      770 GGCAGGCAAGCGGATTTTGGTCCCGGGTGGGGAAGCAAGCTGTGCGAGTGAACAGCT 829
Db      181 ggcagcgaagcgagcttcttggcccggtctgcaacagagctctgtgaaagacagct 240
QY      830 GCTGATCTGTGCGCATCCGGAGTGGGTAGTGAAGAGTCCCTGCGAGCTGGAGCCCAA 889
Db      241 cctgactcttctgctgagctcgagctcgagcgagagagcttcaacagctagaagacaa 300
QY      890 GCGCCAAAGCTGAACACAGCAGATCGCCAACTGATGGAATGGAATGGAAGCTGTAAAGA 949
Db      301 gcggcagagagctgaaacagcagagatggcaagctaaactcaagatgaagctgtaagaa 360
QY      950 GAGCATTTAGAAGCGCAAGGAGAGACTCTCCAGCTCAAGATGTCATCAGCCAGACCGA 1009
Db      361 gagatagaagaacccaagcagagactctgacgtcaagaatgcatlaagcagaagaa 420
QY      1010 GCATTCTCTACAAAGAGCTCATGGCCCAAGACCAAGCTGTCCTGCCATCCGACCT 1069
Db      421 gcaactctacaagaagctgtagtgcacgaacacccaactgcaactgcccacatcgct 480
QY      1070 GCTCCCAAGAGAAGAGATGCGGCTCCCTCCCGCAAGGCGCACTGGGGCTCCGCGCT 1129
Db      481 gctccctgagaagatgtagcgtgcttccaccccccaagtaactcgtgggttcgcgct 540
QY      1130 ACACAATGCTTTGATTAATTCCTGTGCGCTCACTGCTGCTTCCCGGTCTACGTCCTA 1189
Db      541 acacaactgcttgcattactctgtgcccctgtagcgtctgcttctctcgtccta 600
QY      1190 TGACAGTACACAGTTGCTTTGGACACTACCTGATCCCTGGTCAAGAGGCTTTTCA 1249
Db      601 tgacagtgacagtttgccttggagctacacgtgaccccttgacagaagagctttca 660
QY      1250 GCGGACAGCAGAGCTTACGTTATGTTACAGAAATGACAGATCGCTTTCAGT 1309
Db      661 ggcacagtagagccaagctttagttagaagaatgagccatcgccgctgtatgt 720
QY      1310 GATACTAGTGGGAGATGAGAGAGCCCGTGTGCTGCGGCTGTGAGTGGAGAGCA 1369
Db      721 ggtgtagtgggagtagtaaaagccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
QY      1370 GTTGATTTCCCTCCACACTGGGAGAGTGAACAAACATGATGATCAATCATGTC 1429
Db      781 gctgcatctcttgcacactcggagagagagagagagagagagagagagagagagagagag 840
QY      1430 AGCTAGTCAATACAGACAGACCTTCTATAACGTCAGTACTGCGGCTGAGTGGC 1489
Db      841 ccggaagtcagacacaaattactgtatcaatgtcagtaagagcgcgggcatgtgtg 900
QY      1490 CCAAGTCCACTTCTACACTGTCCAGTACAGACTGCGCTTGACTTGTGTATCACCGCT 1549

```















Db	2128	aagttatbgygcctgcatacaggctcttcgcgatagfcaagagcttgaggagaccttcg	2187
Oy	2535	AAGGAATTTCAAGCAGGCGCTTGGAGGCAATGTTCCCGAGACAGTTCACGGTGTATG	2594
Db	2188	aagaaatcgcgcgtacacgcggtgagcaatctacccccggagcaatctacacattctag	2247
Oy	2595	TTGACTTATAGCGCGGAGGAAAGTCGTTATGAACATCTTTTGTAGAGGCGTGAATGGCTCC	2654
Db	2248	ttaacctcaagaagaagaacaagccctccatgattcaatcttgagcagactgatagtctgcg	2307
Oy	2655	TACCTGAACAAGGCTGCTGTGTGTGGAAATTCGCCAGGCTCCATTCAGAGAGCCCTTGTG	2714
Db	2308	tatctgcacaagaatcgtggtgtgttggaactcaacggaacacacattgattgactaaag	2367
Oy	2715	TGGCGTCACATTTGGCGCTTCCCATCATAGTGTTGCCGTTACTGAGAGAACAGTTTGAACAC	2774
Db	2368	tgcgcgagatattggaattacccagcttcgcctcttcgcgcctccaggaattctgtaaacaac	2427
Oy	2775	CGAATTTTACCCTCGGAGTGAATTTGAGACAGAGGCGCATCCTGTCATTTGATGACGATGCT	2834
Db	2428	cgattctcgcctcttattgcatgcatggaacagagacgctctctctctgcgcgaacagatgc	2487
Oy	2835	CACCTCCGCGCATGACGAAATCATGTTTGGGCTTCGGGGTGTGAGAGAGATCTCGGAGCGC	2894
Db	2488	catctgcgtcatgtagatctctctctcgcgattccgltgtttgcgcggagacatcgcgataga	2547
Oy	2895	ATGTGGGCTTCCCTGGCGGCTTACACGCGATGGGA-----CATCCCGCATAGTCTCGG	2948
Db	2548	gttgtagagctctcccaagagacgtttacacgcgtttggaatttggaaatcccaatgycagtg	2607
Oy	2949	CTGTACACATCCCAACATACGTCGTGTGAGCGTCCAGTGGTGTGACAGGCGTCCCTCTTT	3008
Db	2608	cactacaactcccaattacacgcctgcgcgaactgagacatggtgtgcgcgaaggagctgcgttga	2667
Oy	3009	CACAAGTATTATGCGTACCTGTAATTTTATTTGATGCGCCCGACGGCCATCCGGAGATGGTG	3068
Db	2668	cacaagttactatcttaactctgtaaacctacacatctgcgcgagcgagataagagacaagtg	2727
Oy	3069	GATGAAATACATCACTGTGAGAGACATTTGCCATGAACCTTGTGTCGCCACATCACTGGG	3128
Db	2728	gattgagtaactagtaactgttggagacatctgcacgtaaattcttcgtatcatgcatataactaga	2787
Oy	3129	AACCCCCCATCAAGGTGAGCTCAGGGTGGAGATTCGATCGATCCGACAGATGGCCCTCAGGCC	3188
Db	2788	aaaccccaagtaaaatctccacccacggttggaaatcttcgcgtccgcgtgtctcgtttca	2847
Oy	3189	CTGTCTCATGATGACATCCCACTTCCACAGAGCGGACACAAGTGCATCAACTTTTCTGTGAAG	3248
Db	2848	ctcagcgagagtagaacctactcttcccaagagcggcgcacaagtgtcatcaattctcttaagccg	2907
Oy	3249	GTTTACGGCTACATGCGCCTCTCTGTACACGACGATTCAAGGTGAGATTCTGTGCTTTCAAG	3308
Db	2908	gtttctggtctatacgccggtctgtgacaacccagtaaccgagcgagatccatctcttcaag	2967
Oy	3309	ACACGCGTCCGCCATGACGAAGCAAGTGTGTAAGTTCACTCATGAGGGGCA	3368
Db	2968	acgcggaatccgcgcagacaagaanaagtgcttcaagtaactacatctgcatgta 3017	

RESULT	9
ABL26682/c	
ID	ABL26682 standard; DNA; 6105 BP.
XX	
XX	
AC	ABL26682;
XX	
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 31519.
XX	
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ds.	
XX	
XX	

OS	Drosophila melanogaster.	
PN	WO200171042-A2.	
PD	27-SEP-2001.	
PF	23-MAR-2001; 2001WO-US09231.	
XX		
XX	23-MAR-2000; 2000US-191637P.	
PR	11-JUL-2000; 2000US-0614150.	
PA	(PERE ) PE CORP NY.	
PI	Venter JC, Adams M, Li PWD, Myers EW;	
DR	WPI; 2001-656860/75.	
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	
PT		
XX		
PS	Claim 1; SEQ ID NO 31519; 21pp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57373-ABB72072).	
CC		
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.	
CC		
XX	Sequence 6105 BP; 1624 A; 1403 C; 1377 G; 1701 T; 0 other:	
SQ		
Query Match                      8.8%; Score 545.6; DB 23; Length 6105; Best Local Similarity 60.9%; Pred. No. 2.1e-132; Matches 944; Conservative 0; Mismatches 594; Indels 12; Gaps 3		
OY	1618 CCGACTAGTGGGCACTGTGGAGAGCGGAGACCGCTTGGAATTGCTGAGCTGCC 1877   DB 3125 CTGCACAGATTTGGAGCGTATGTGGATTTCATCAGCGCCGCAATTTGCTTAAGAGACTCT 3066   OY 1878 ACCCTGGCCCATCATATACCACCGGGGAGCCCTGCTGTAATTTCCCTGGGGTGTGA 1937   DB 3065 ACGTTTTCTCGAATTTTACCCCCGTTGAA--TGACAGATTTTCGTGCACACTAATGCTT 3009   OY 1938 ACACGGCTCTTCGAAAGCCCTGGAAGTGCCTGCCCTCCGCTGGTCTGTGG--GGAGCAG 1994   DB 3008 GCTAGAGATTTTGAAGAAGCTCTCCGTTCCGGAGACAGTACTGTTATCCTTGGACAATATGAG 2949   OY 1995 GTCCAGCTTCCCTAACAGAGACATGCTGCAGTGGAAAGAGGCGCCCTGTGGTGCATAAG 2054   DB 2948 TTGGTTTGGCGGTATGTGGAACATGTAGATTGGAGAAAGAGCGCACTACTGTACCGAAA 2889   OY 2055 CCTCGTATTCCGAGGATTCATTTCTGTGTGCAGAAAGCCCTCCGATAGTAGACTCTGTGCT 2114   DB 2888 GCTCGATCACAGAAATGCAATTTCTTCTGTGGCGGAGTGACAGATGCGGACTTAATGCTG 2829   OY 2115 ATGAGGGGAGCAAGACCGCTTCTCTGGAGAGACTTACTTCTCCACTGTGACAGTATTTT 2174   DB 2828 CTGGAGACGACAGGGGCGACATCATTTGGGAACCGCTAATTTTGAAGCTTGTGCAGGCCACCGTG 2769   OY 2175 AATACGCTGTGGCTATGATTAGAGACTTCGCATCCAGATCCAGGCGCTCCCATCCGGGAA 2234   DB 2768 GATACAGTATTTGCCAGATTNAAAGGATTCGCTTGGAAATCCGGCCCAAGACCGGTGCGCTCG 2709   OY 2235 GAGGCGGACAGTGCAGATCCCACCGTTTCAGGCAAGGCGGCTGAGACTGACCCCAATG 2294   DB 2708 GTAATACGCAAAAGTGTGTTCAACAGACATTTATACCCCTAAATCCGATCCACCTGTT 2649	





D 239 AGTGCATCTGAGAGACAGCCCTGGAGGTCTCCTTTCAAGAACTCCGTGCTGAGG 180  
Q 73 TGTATGCTACACAAAGTCAGAGAGAGGCTCTGAACACATGCGCTGATTGTTGCC 132  
D 179 TGTATGCTACACAAAGTCAGAGAGAGGCTCTGAACACATGCGCTGATTGTTGCC 120  
Q 133 AAAGGATCATAGAACTGGCATTTATTCTGTTCTTAACCTATTACTGTTAACTGTA 192  
D 119 AAAGGATCATAGAACTGGCATTTATTCTGTTCTTAACCTATTACTGTTAACTGTA 60  
Q 193 ATAGACACTATGATTTGTTGGTCAGAAACCAAGAA 232  
D 59 ATAGACACTATGATTTGTTGGTCAGAAACCAAGTA 20  
RESULT 13  
AAF21849  
AAF21849 standard; DNA: 3008 BP.  
AAF21849;  
27-MAR-2001 (first entry)  
Human breast and ovarian cancer associated antigen gene SEQ ID 236.  
Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
antidiabetic; antiinflammatory; antileuk; vulnery; anticonvulsant;  
antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
Addison's disease; allergy; autoimmune haemolytic anaemia;  
autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
cardiovascular disorder; wound healing; neurological disease; ds.  
Homo sapiens.  
X OS  
X PN NC020005173-A1.  
X PD 21-SEP-2000.  
X PF 08-MAR-2000: 2000MO-US05881.  
X PR 12-MAR-1999: 99US-0124270.  
X RA (HUMA-) HUMAN GENOME SCI INC.  
X Rosen CA, Ruben SM;  
X WPI; 2000-611515/58.  
X P-PSDB: AAB58946.  
X DR  
X DR  
X PT New human breast and ovarian cancer associated gene sequences and the  
X PT polypeptides encoded by these genes, useful in the prevention,  
X PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
X PT disorders and neurological diseases -  
X PS  
X PS Claim 1; Page 660-661; 1299pp; English.  
X CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
X CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
X CC associated with breast and ovarian cancer. Included in the invention are  
X CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
X CC isolation and characterisation of the DNA and protein sequences of the  
X CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
X CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
X CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
X CC antidiabetic; antiinflammatory; antileuk; vulnery; anticonvulsant;  
X CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
X CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
X CC particularly breast and ovarian cancer. The nucleic acid sequences,  
X CC proteins, agonists and antagonists may also be used in the diagnosis,  
X CC prevention and treatment of immune disorders e.g. Addison's disease,  
X CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.  
X CC  
X CC  
X CC Sequence 3008 BP; 792 A; 711 C; 737 G; 765 T; 3 other;  
Query Match 3.3%; Score 200.6; DB 21; Length 3008;  
Best Local Similarity 58.0%; Pred. No. 8e-42;  
Matches 397; Conservative 0; Mismatches 279; Indels 9; Gaps 2;  
Q 2668 TCGTGGTGTGTGGAATTTCCCAAGCTGCCATCAGAGACCTTCTGTGCGCTGACATTG 2727  
D 1596 ttgtcgtcgtggaataacagaataaaacccccaagaattctctcgtcccaaaatcc 1655  
Q 2728 GCGTTCATCATGAGTGGTCCGTGCTGAGAAACACTTTGAACACCGATTCTTACCGT 2787  
D 1656 ggggtccattaaagttcgtgagacgtcgtgaaacaaagttaaagtaaccgttcttccct 1715  
Q 2788 GGAATGAATTTGAGACAGAGGCCATCCTGTCATTTGATGACGAT--GCTCACCTCCGCC 2844  
D 1716 atgtatgaatcagagacagaagctgtctgtccatctgtatgatatacatatgtgcgcct 1775  
Q 2845 ATGACGAATCATGTTTGGTCCGGTGTGAGAGAACTCGGAGCCGATCGTGAGCT 2904  
D 1776 ctgagcagctcgaatctgtatgagttcgtgcggaattctccagccggttggtagt 1835  
Q 2905 TCCCTGGCCGTTTACACGACATGGAGCATCCGCCATCAGTCCGTGCTTACAACTCCAACT 2964  
D 1836 acccggtgtcgtcagcctcctcgtgagacataagatgaatgaagttgagtcgtgagt 1895  
Q 2965 ACTGCTGTGAGCTGTCCATGATGCTGTGACAGAGTGTGCTGCTTCTTTCACAACTATTAGCT 3024  
D 1896 ggaagaaatgaagttcctcatgtgtcactgtgagcagcttcttaccagaatattatc 1955  
Q 3025 ACCTGTATTCTTATGATGATGCCCGACGCATCCGGACATGCTGATGATGATCACT 3084  
D 1956 accgttatacttaaaataatcctgtgagatatacagaactgtgtgagatcgtcatatgact 2015  
Q 3085 GTGAGCAATTTGCCATGACATCTGCTGTCACAGAGTGTGCTGCTTCTTTCACAACTATTAGCT 3144  
D 2016 gtgaaatattgtccatgaactctcgtgccaagtcgaaggaagaaagcagttatcaagg 2075  
Q 3145 TGACCTCAGCGTGGACATTCGATGCCAGATGCC-----TCAGCCCTGTCTCATG 3198  
D 2076 taaccaccaagaagaataatcaagtgtcctgtgacagacatagatgaggtcttcaactag 2135  
Q 3199 ATGACTCCCACTTCCACAGAGCGGACAAAGTGCATCAACTTCTTGTGTAAGGTGTACGGCT 3258  
D 2136 accaaacacacatgtgtgagagtcagagatgcatacaagaattgtcttccagttcggga 2195  
Q 3259 ACATGCCCTTCCTGTACACGACGATTCAGGGTGTGCTGTGCTTCAAGACACGCGCTGC 3318  
D 2196 ccatgcctctcaaggtgtgtggaacacagcgtgacccctgtcctgtacaaagatgacttcc 2255  
Q 3319 CCCATGACAAGACCAAGTGTCTTCAA 3343  
D 2256 ctgagaagctgagaagcttcccca 2280  
RESULT 14  
AA208224  
ID AA208224 standard; cDNA: 3175 BP.  
X AC AA208224;  
X X  
X X 25-JAN-2000 (first entry)  
X DT  
X X  
X X Human Ext-2 cDNA.  
X DE Human Ext-2 cDNA.  
X X  
X X Human Ext-2 cDNA; Ext-2 protein; chromosomal position 11p11-13;  
X KW hedgehog protein; hedgehog interacting protein; diffusion regulation;

KW glycosyltransferase; GAG chain; protein core; proteoglycan;  
KW hedgehog-mediated signal transduction; hedgehog specific GAG chain;  
KW extracellular matrix; drug discovery assay; growth factor; apoptosis;  
KW cell proliferation; cell differentiation; cancer; ds.

OS Homo sapiens.

Key	Location/Qualifiers
FT CDS	335..2491
FT	/*tag= a
FT	/note= "Human Ext-2 protein"

PN W09950385-A2.

PD 07-OCT-1999.

PF 30-MAR-1999; 99WO-US06892.

PR 30-MAR-1998; 98US-0079928.

PA (HARD ) HARVARD COLLEGE

PI Bellaiche Y, The SI, Perrimon N;

WPI; 1999-610841/52.

DR P-PSDB; AAY28883.

PT Identifying inhibitors of signal transduction of extracellular proteins, used for developing agents for treating, e.g. proliferative disorders or cancers or for cartilage -

Example 3; Page 99-103; 141pp; English.

The present sequence encodes for human Ext-2 protein which is the putative tumour suppressor for multiple exostoses. This gene is localised to chromosomal position 11p11-13. Exts are hedgehog interacting proteins which bind to hedgehog proteins with high affinity to regulate diffusion. They are a family of glycosyltransferases synthesising GAG chains attached to the protein core of proteoglycans. Hedgehog-mediated signal transduction can be inhibited by inhibiting the synthesis of hedgehog specific GAG chains by Ext dependent pathways. Drug discovery assays identifies such agents that modify the biological activity of growth factors by inhibiting Ext activity to prevent hedgehog-dependent proliferation of cells. This is useful in treating disorders related to aberrant apoptosis, cell proliferation or differentiation, cancer etc.

Sequence 3175 BP; 760 A; 792 C; 817 G; 786 T; 0 other;

Sequence 3175 BP; 780 A; 792 C; 817 G; 786 T; 0 other;

Query Match	3.38;	Score 200.6;	DB 20;	Length 3175;
-------------	-------	--------------	--------	--------------

Best Local Similarity 58.0%; Pred. No. 8.2e-42;  
Matches 397; Conservation 0; Misses 370

Matches	391;	Conservative	0;	Mismatches	279;	Indels	9;	Gaps	2;
---------	------	--------------	----	------------	------	--------	----	------	----

[illegible]

QY	2965	ACTCTGTGACACTTCATCGATGGTGTGACAGTGTCTGCTCTTTCACAGATTTATGCT	3024
Db	2091	ggagcaatgaagatgctccatggtgctcactcctgggagcctttatccaagatattlaatt	2150
QY	3025	ACCTGTATTCTTATGTGATGCCAGGCCATCCGGAGCATGGTGATGATACATCAACT	3084
Db	2151	acctgtatacctatacaaatgctcgtgggatatcaagaactcggtagatgtctatagaact	2210
QY	3085	GTGAGAGACATTGGCATGAACTTCTTGTCTCCACATCACTGGAAACCCCATCAAGG	3144
Db	2211	gtgaagatatgtgccatgaactctctgttgcacagctcacggaaagaacgattacaagg	2270
QY	3145	TGACATACGGGTGACATTCGATTCGCCAGAGATGCC-----TCAGGCCGTCTCATG	3198
Db	2271	taaccccacgaagaataattcaagtctccgtgagtgccagcatagatggcttaccag	2330
QY	3199	ATGACTCCCATCTTCACAGACGGGCACAGATGCATCAACTTCTTCCTGAAGGTATACGGT	3258
Db	2331	accaaacacacatggtggagatgctcaagatgatacaacaagtttcttcagcttcttggga	2390
QY	3259	ACATGCCCCCTCTCTGTACACGCAAGTTACAGGTGAGATTCTGTCTTTCAGAACACGCTGC	3318
Db	2391	ccatgacctcttaagaatgtgtgaaacacccgagctgaccccttctctgfacaaagaatgatcttc	2450
QY	3319	CCCATGACAAAGCAACAGTCTTCAA	3343
Db	2451	ctgagaagctgaagaagcttcccaa	2475

RESULT 15

ID	AAV19375 standard; CDNA; 3003 BP.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

AC AAV19375

DT 20-AUG-1998 (first entry)

DE Hereditary multiple exostose associated EXT2 gene isoform encoding cDNA.

KW Hereditary multiple exostose; EX12; chondrosarcoma; human; isoform  
KW treatment; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
100		100
101		101
102		102
103		103
104		104
105		105
106		106
107		107
108		108
109		109
110		110
111		111
112		112
113		113
114		114
115		115
116		116
117		117
118		118
119		119
120		120
121		121
122		122
123		123
124		124
125		125
126		126
127		127
128		128
129		129
130		130
131		131
132		132
133		133
134		134
135		135
136		136
137		137
138		138
139		139
140		140
141		141
142		142
143		143
144		144
145		145
146		146
147		147
148		148
149		149
150		150
151		151
152		152
153		153
154		154
155		155
156		156
157		157
158		158
159		159
160		160
161		161
162		162
163		163
164		164
165		165
166		166
167		167
168		168
169		169
170		170
171		171
172		172
173		173
174		174
175		175
176		176
177		177
178		178
179		179
180		180
181		181
182		182
183		183
184		184
185		185
186		186
187		187
188		188
189		189
190		190
191		191
192		192
193		193
194		194
195		195
196		196
197		197
198		198
199		199
200		200
201		201
202		202
203		203
204		204
205		205
206		206
207		207
208		208
209		209
210		210
211		211
212		212
213		213
214		214
215		215
216		216
217		217
218		218
219		219
220		220
221		221
222		222
223		223
224		224
225		225
226		226
227		227
228		228
229		229
230		230
231		231
232		232
233		233
234		234

```
/*tag= a
```

PN EP837127-A2.

PD 22-APR-1998.

PF 26-AUG-1997; 97EP-0306503.

PR 21-OCT-1996; 96CN-0121928.

PA (UYHU-) UNIV HUNAN MEDICAL.

PI Deng HX, Fan CH, Xia J,

DR WPI; 1998-219110/20.

XX

PT exostoses or chondrosarcoma

PS Claim 4; Pages 28-31; 31pp; English.

CC This cDNA encodes an isoform of the EXT2 gene associated with hereditary  
CC multiple exostoses and chondrosarcoma. The polynucleotide is an isoform  
CC of the EXT2 gene described in Nature Genet., 14, 25, 1996. The  
CC polynucleotide can be used in the detection and treatment of EXT2-related





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:01 ; Search time 26.04 Seconds  
(without alignments)  
3391.171 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873  
Sequence: 1 MTGYTLNRNGAGNGGQTCM.....DSVLEKTRLPDHTKCFKFI 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Database: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4873	100.0	919	2 JC5934	exostose-related p
2	1252.5	25.7	814	2 T23200	hypothetical prote
3	367	7.5	330	2 JC5935	exostose-related p
4	360	7.4	334	2 T49195	hypothetical prote
5	306.5	6.3	764	2 T48446	hypothetical prote
6	288	5.9	329	2 D96834	hypothetical prote
7	232.5	4.8	444	2 T20803	hypothetical prote
8	123.5	2.5	425	2 T05631	hypothetical prote
9	121.5	2.5	511	2 T51544	hypothetical prote
10	120.5	2.5	680	2 AG2419	hypothetical prote
11	120.5	2.5	1202	2 SS5553	hypothetical prote
12	120	2.5	460	2 F84764	LAR-interacting pr
13	120	2.5	893	2 F72253	hypothetical prote
14	119.5	2.5	824	2 I50618	c-fps proto oncoge
15	118.5	2.4	622	2 S45129	VPS27 protein - ye
16	117	2.4	440	2 C96697	hypothetical prote
17	117	2.4	440	2 T02165	hypothetical prote
18	115.5	2.4	1331	2 A72647	probable surface 1
19	115.5	2.4	811	2 JC7619	hypoxia-inducible
20	114.5	2.3	1024	2 G71434	probable limonene
21	114	2.3	670	2 T38446	microtubule-associ
22	113.5	2.3	385	2 H83930	lipase (esterase)
23	113.5	2.3	359	2 S74668	hypothetical prote
24	113	2.3	676	2 F69394	H+-transporting Ar
25	113	2.3	1637	2 AE2109	two-component hybr
26	113	2.3	2663	1 S28261	centromere protein
27	112.5	2.3	4464	2 D87755	protein T21E12.4
28	111.5	2.3	2484	2 T26216	hypothetical prote
29	111.5	2.3	2607	2 T26215	hypothetical prote

30	111.5	2.3	3075	2 S14458	lamin alpha-1 ch
31	110.5	2.3	382	1 I40209	cytochrome p450 B7
32	110.5	2.3	1940	2 A29320	myosin heavy chain
33	110.5	2.3	5138	2 B96695	hypothetical prote
34	110	2.3	1207	2 T13827	kinesin-73 - fruit
35	109.5	2.2	1207	2 B88789	protein ZK1251.9
36	109.5	2.2	1596	2 A41216	guanine nucleotide
37	109	2.2	475	2 T47480	hypothetical prote
38	109	2.2	765	2 E96558	hypothetical prote
39	109	2.2	978	2 A70387	conserved hypocher
40	109	2.2	4967	2 S72269	ryanodine receptor
41	108.5	2.2	346	2 T48520	hypothetical prote
42	108.5	2.2	779	2 AG1978	hypothetical prote
43	108.5	2.2	1212	2 T00332	hypothetical prote
44	108	2.2	720	2 D98141	hmel2 protein (Af0
45	108	2.2	720	2 AF3146	polysaccharide blo

## ALIGNMENTS

Query Match	Best Local Similarity	100.0%	Score 4873	DB 2	Length 919	Matches 919	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	MTGYTLNRNGAGNGGQTCMLRMSNRIRLTWLSPTLVILVFPPLIAHYVLTTLDEADEA 60									
Db 1	MTGYTLNRNGAGNGGQTCMLRMSNRIRLTWLSPTLVILVFPPLIAHYVLTTLDEADEA 60									
QY 61	GKRIFGRVGNELCEVKAHVLDLCIRISVSPELLQLEAKROELNSEIATKMLTKTEACKS 120									
Db 61	GKRIFGRVGNELCEVKAHVLDLCIRISVSPELLQLEAKROELNSEIATKMLTKTEACKS 120									
QY 121	IENKODLQKNTVISOETHSYKELMAONOPKLSLPIRLLPKODAGLPPKATRGCRRLH 180									
Db 121	IENKODLQKNTVISOETHSYKELMAONOPKLSLPIRLLPKODAGLPPKATRGCRRLH 180									
QY 181	NCFDYSRCPITSGPVPVYVSDQFVFGSYLDPVKAQFOATARAVVYVENADIACLYVI 240									
Db 181	NCFDYSRCPITSGPVPVYVSDQFVFGSYLDPVKAQFOATARAVVYVENADIACLYVI 240									
QY 241	LVGMOEPVYLPAELKOLYSLPHMRTDGHNVITLMSKSDTQNLINVTSTRANVAQ 300									
Db 241	LVGMOEPVYLPAELKOLYSLPHMRTDGHNVITLMSKSDTQNLINVTSTRANVAQ 300									
QY 301	STFTVYVRGPFDLVPSPLVHAMSEPMETPPQVPKRYLPTFOGEKTESLRSSLOEA 360									
Db 301	STFTVYVRGPFDLVPSPLVHAMSEPMETPPQVPKRYLPTFOGEKTESLRSSLOEA 360									
QY 361	RSFEEMEGDPADYDRRIATLAKAVDSKLDQVLEFTCKNOPKPSLPTEMALCGERED 420									
Db 361	RSFEEMEGDPADYDRRIATLAKAVDSKLDQVLEFTCKNOPKPSLPTEMALCGERED 420									
QY 421	RLELKLTSTALITTPGDPRLVISGCAITRLEALEVGAAPVVLGEOVLPIYQDMLOWNE 480									
Db 421	RLELKLTSTALITTPGDPRLVISGCAITRLEALEVGAAPVVLGEOVLPIYQDMLOWNE 480									

Db 421 RLLKLKSTFALLITIRGDPRLVISSGCATRLFEALVEGAVPVVLGEVOLPYQDMLOWNE 460

QY 481 AALVVPKPRVTEVHFLLRSLSDSDLLAMRKGRLMETYSTADSIENTVLAMIIRTRIQI 540

Db 481 AALVVPKPRVTEVHFLLRSLSDSDLLAMRKGRLMETYSTADSIENTVLAMIIRTRIQI 540

QY 541 PAAPFREAAAEIPHRSGKAGTDPNNADNGDLDLGVETEPYPASPRYLINFTLVYTF 600

Db 541 PAAPFREAAAEIPHRSGKAGTDPNNADNGDLDLGVETEPYPASPRYLINFTLVYTF 600

QY 601 YRSNMCAGPFLHPHTPFDVLPVLPSEAKFLTSGTGFRRPFGGAGSGSKETQALGQVNR 660

Db 601 YRSNMCAGPFLHPHTPFDVLPVLPSEAKFLTSGTGFRRPFGGAGSGSKETQALGQVNR 660

QY 661 EGFVYVMLITYREXYLAMSRLNGPLPLKVVYVWNSPKLPSEDLLMPDIGVIMVVRT 720

Db 661 EGFVYVMLITYREXYLAMSRLNGPLPLKVVYVWNSPKLPSEDLLMPDIGVIMVVRT 720

QY 721 EKNSLNRRFLPMNETETETALISIDDDAHLRHDEINFGFRVREARDRIVGPGRYHAMDI 780

Db 721 EKNSLNRRFLPMNETETETALISIDDDAHLRHDEINFGFRVREARDRIVGPGRYHAMDI 780

QY 781 PHQSWLYNSNTSCELISWLTGAAPFHKKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFTLV 840

Db 781 PHQSWLYNSNTSCELISWLTGAAPFHKKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFTLV 840

QY 841 SHITKRPPIKXTSRMTFPCPCCPALSHDSDHFERHKCNINFTKYGYGMPLLYTGFRD 900

Db 841 SHITKRPPIKXTSRMTFPCPCCPALSHDSDHFERHKCNINFTKYGYGMPLLYTGFRD 900

QY 901 SVLFKTRLPDHDKTCFKFI 919

Db 901 SVLFKTRLPDHDKTCFKFI 919

RESULT 2

T23200

hypothetical protein K01G5.6 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 15-Sep-2000

C:Accession: T23200

R:Wall, M.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19706

A:Accession: T23200

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-814 <MIL>

A:Cross-references: EMBL:Z92803; PIDN:CAB07245.1; GSPDB:GN00021; CESP:K01G5.6

A:Experimental source: clone K01G5

A:Genetics:

A:Gene: CESP:K01G5.6

A:Map position: 3

A:Intons: 7/1; 52/2; 78/2; 112/1; 156/2; 216/3; 702/3; 753/2

C:Superfamily: *Caenorhabditis elegans* hypothetical protein K01G5.6

```

Query Match          25.7%; Score 1252.5; DB 2; Length 814;
Best Local Similarity 33.4%; Pred. No. 5.5e-82;
Matches 313; Conservative 147; Mismatches 312; Indels 165; Gaps 25;

QY      21  LRMSRIRLITLMSFLPFLILWFEPL--TAHYLYL-----TLDEADAGAKRI 64
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       5  LMGSSRSFVPSLRVSAFLIFEFVYTYIIIVNSSEPSWITQDALKONIEMLDYDASC 64

QY      65  FGPRIANGNEICEKVAHLDLCRIRESYE--ELIOLFAKROELNSEIAKLNKIFACKSTEN 123
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      65  SGYSIGRIILREQKRIILASVRLLETITSQYKIEIRIVQEBLQRIIPQOKOLEISALGCTEA 124

QY     124  AKODILQLKNVISQTEHSYKELMAONQPKLPIRIILPEKDDAGLPPPKATRGCRILHNCF 183
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     125  AQRQL-----EELRTQNQNVKVFIPFS--FLQIPRELDQPSQISPNQLDIT 168

```

[illegible]

```

RESULT      3
CS5935
Xsotose-related protein 2 - human
Species: Homo sapiens (man)
Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
Accession: J05935
Saito, T.; Seki, N.; Yamauchi, M.; Tsuji, S.; Hayashi, A.; Kozuma, S.; Hori, T.
Biochem. Biophys. Res. Commun. 243, 61-66, 1998
Title: Structure, chromosomal location, and expression profile of EXTR1 and EXTR2
Reference number: J05934; MUID:98139867
Accession: J05935
Status: preliminary
Molecule type: DNA
Residues: 1-330 <SAI>
Cross-references: DDBJ:AB009284; NID:g27233392; PIDN:BA024081.1; PID:g27233393

Query Match      7.58; Score 367; DB 2; Length 330;
Best Local Similarity 30.88; Pred. No. 1e-18;
Matches 86; Conservative 49; Mismatches 98; Indels 46; Gaps 7;

```

```

OY 656 GAVPEOEFVWVLATEREVLNMSLERLNGLPYLNKVVVVVMSKPLSDEDLDPDGG--- 712
Dd 59 GKSTMSDFLLIMOTNRNRDLDLLKLNHYQAVPRLNHVIYVWNNGIEKAPDELNNSLGGPR 118
OY 713 VPIVWVTEKNSLNNRFLPEWNEIETFEALISIDDAAHLRDEIMFGFRVWRREARDIVGFP 772
Dd 119 IPVIRKOOTANMRMRRLQVFPPELFTNAVLMVDDDLTISTPDLFAFASVMOQFPDOJGVFV 178
OY 773 GRVHAMDPHOSMLVN-----SNVSCELSVLTGAAFPKRYAYLYSVVMOAIR 822
Dd 179 PRKH---VSTSGITSYGSGFEMQAPGSGNGDOYSWVLIGASFNSKYLELQF--QPAVH 234
OY 823 DWAVDEYINCEDIAMNFLVS-HITRKPIKV-----TSRWTFRCPCGPQA 865
Dd 235 ALIDTQNCDDIAMNFIILKHIGTKTSGIEFVKPVNMNMLEKETNSGSGMW----- 284
OY 866 LSHDDSHFERHKCNFVKVGYGVMLLYTOFRVDSVLF 904
Dd 285 --HRAEHALORSTCNKLNVNITDSDMPLRYSNIMISQFCF 321

```

RESULT 4  
T49195  
hypothetical protein F27K19.10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49195  
R:Benes, V.; Wurmbach, E.; Dzonek, H.; Ansorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.,  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25014  
A:Accession: T49195  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-334 <BEND>  
A:Cross-references: EMBL:AL163832; GSPDB:GN00061; ATSP:F27K19.10  
A:Experimental source: cultivar Columbia; BAC clone F27K19  
C:Genetics:  
A:Gene: ATSP:F27K19.10  
A:Map position: 3  
A:Introns: 71/2; 203/3; 258/2; 282/1

[illegible]

RESULT 5  
T48446  
hypothetical protein T32M21.100 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48446  
R:Beran, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.;  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
Submitted to the Protein Sequence Database, March 2000  
A:Reference number: 224487  
A:Accession: T48446  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-764 <BEV>  
A:Cross-references: EMBL:AL162875  
C:Experimental source: cultivar Columbia; BAC clone T32M21  
C:Genetics:  
A:Map position: 5  
A:Introns: 105/3; 146/3; 202/3  
A:Note: T32M21.100

[illegible]

RESULT 6  
D96834  
hypothetical protein F516.4 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96834  
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alon  
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hutzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talia  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141, MUID:21016719  
A:Accession: D96834  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <STO>  
A:Cross-references: GB:AE005173; NID:g6751704; PID:ANF27686.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F516.4  
A:Map position: 1

Query Match	5.9%;	Score 288;	DB 2;	Length 329;
Best Local Similarity	27.7%;	Pred. NO. 5e-13;		
Matches	79;	Conservative	52;	Mismatches 106;
			Indels	48;
			Gaps	8;





Db 538 MADGHTD-----SYSTSAVLRPRRQKRLAALRDEPSKVOTLNEQDMERAQASVLANVAQ 592  
Qy 454 ALEVAVPVVLC-----QVQLPYQDMLQWNEALVVPKPRVTEVHFLRLSLSDSLLAMR 510  
Db 593 APESDA-DVSDGDDRDTLLSSVDLLSPSGNA-----DAHTLAMLQe-QDAINK 641  
Qy 511 QGRFMEYFST---ADSIENFV-----LAMIRTRIQIPAPPIREFAA----- 551  
Db 642 EIRLQIEKEKENTQRAEELIESRVSGSLDNIGRFRSMSSIPYPASSSLASSSPSGGRST 701  
Qy 552 --EIPH-----RSGKAAGTDPMNAD-NGDLDLGPVETEPYPASRYLRNFTL----- 595  
Db 702 PRRIHSPAREVDRLGVMTLLPSPREVRDCKTITKCESPP-SSPRALRLDLRLKGAH 760  
Qy 596 -----TVDFPRSNWCACGP-----FHLPHFPDPVLPSEAKFLGSGTGR 637  
Db 761 TVSHEDIDIRNLTGSODGPVSNPSSNSQDSLHKAP-----KKKQIKSSIG-R 809  
Qy 638 PIG-----GAGGSGKEF--QAA-----LGGNVPRQFTVVMILTYEREDEV 675  
Db 810 LFGKKERKRPQOTGREALGQAGVSETDNSQDLAGSLKLGQAEKNR-----KIQKHEH 864  
Qy 676 LMSLERLNGLPYL--NKVYVYVNSPKLPSEDLMPDIPGVIMVYRTEKSLNRFPLPW 732  
Db 865 LEEA--RROGLPFAQWDGPVYVWLE-----LM-VGMPAWVYACRAN----- 904  
Qy 733 NEITEALSTDDAHLRHD 752  
Db 905 --VKSALMSALSTPEIQE 922

RESULT 12  
F84764  
hypothetical protein Atg35100 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cross)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: F84764  
R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487  
A/Accession: F84764  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1460 <STO>  
A/Cross-references: GB:AE002093; NID:g3668093; PIDN:AAC61825.1; GSPDB:GN00139  
C/Genetics:  
A:Gene: Atg35100  
Map position: 2

Query Match 2.5%; Score 120; DB 2; Length 460;  
Best Local Similarity 19.6%; Pred. No. 1.1;  
Matches 105; Conservative 79; Mismatches 162; Indels 190; Gaps 26;

Qy 146 MAONQPKLS-LPRLRLPEKODAGLPKATRGRLHNCFDYSRCLTSGFPVYVD-SDQ 203  
Db 29 VSSRLSPSSDPLRLIREDD-----DEGRAPIDPRVRYVMYNPKR 70  
Qy 204 FVFGSLDPLVKAPQATAAN----- 225  
Db 71 FTYG-----LIEQ--HSIAGGIKKPVGDVYTLKYPGHMHBMVLFSDLNQEPVDRSGS 123  
Qy 226 --VYTEMADIACTVYLIVGEMQEPV-VLRAPEL-----EKOLYSLPHM-----RT 268  
Db 124 PIYRVSDPADADLFYVPSVSSSLIYNAGRPRVDAGSGYSDERKQEGILVELEQEMWRRN 183  
Qy 269 DGHNVHILNLSRKSQTONLILYN--STGRAMVAOSTFYTYQYRPGDL--VSPPLVHAM 324  
Db 184 AGHDHYI-----PAGDPNMLYRILDRVKNANVLLVSDFGRLRPQGSFVADVIFPIYSHRYA 238

Qy 325 EPNEMEIPQVPEV-KRYLTFEQGEKIESLRSLQEANSFEEEMEGDPADDRITATL 383  
Db 239 IEN-----GEIGVEDRNTLLFFMGNRYK-----DGKVRDLDFQVL--- 275  
Qy 384 KAVQDSKLDQVLEFTCKNQRPSPSLPTEMALGGERDRELK-LSTGALLTP-GDRL 441  
Db 276 -----EKEDDVITKIGTOS-----RENRAATKGNHTSKFCILNPAGD--- 312  
Qy 442 VISSGCATRLFEALVGAVPVVLGEOVLQYQDMLQWNEALVVPKPRVTEVHFL---LR 498  
Db 313 ---TFSACRLPDSIVSLCPPLVSDSIELPREDVIDYKFSIFVANAALDPGLVQMLR 369  
Qy 499 SLSDSDLLAMRQGRFLMEYTFSTADSIENFVLAIRTRIQIPAPPIREFAAETPRSG 558  
Db 370 KIKTKILEYQREMSKFD-----FTKLMSM-----HK-- 398  
Qy 559 KAAGDPMANNGDLDLGPVETEPYPASRYL--RNFLLTYTDFPRSNWCACGP 612  
Db 399 -----PNRSYFHLNLFVWERPFGLMRSLRREFTEYTSRY-----SPLPMH 443

RESULT 13  
F72253  
hypothetical protein TM1450 - Thermotoga maritima (strain MSB8)  
C/Species: Thermotoga maritima  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C/Accession: F72253  
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A/Reference number: A72200; MUID:99287316  
A/Accession: F72253  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1893 <ARN>  
A/Cross-references: GB:AE001796; GB:AE000512; NID:g4982004; PIDN:AAD36518.1; PID:g498  
A/Experimental source: strain MSB8  
C/Genetics:  
A:Gene: TM1450

Query Match 2.5%; Score 120; DB 2; Length 893;  
Best Local Similarity 19.3%; Pred. No. 3.3;  
Matches 107; Conservative 84; Mismatches 194; Indels 170; Gaps 24;

Qy 247 EPPVLRPALEKQLYSLPHRTDGHNVITNLSRKSQTONLILYN--VSTGRAMVAOST-- 302  
Db 30 EDVVLSPKIKRKRIEL--WRLNGENLKIATTLKALTEKRVFSPDLRENSLMARSTKL 87  
Qy 303 -----FYVQ-----YRPG--FDLVVSPVLVHAMSEPNFEIP 332  
Db 88 TLSPEKIVEMGYELVFTVQVNGEFAIRGLIIDYSPDNPPVRIELFGDEIEIRFFKVD 147  
Qy 333 PQVPVKKRY-----LTFQGEK-----IESLRSLQEARSEEMEG 369  
Db 148 TQ-----RSFGVMDKTLILPVDYGGSTLIDFLKTARFICEDLQVLDYRKRFRKEMRD 203  
Qy 370 DPPADDDRIATLKAQVDSKLDQVLEFTCKNQRPSPSLPTEMALGGERDRELKLT 429  
Db 204 -----LLKRYNDFDERVVEVLKAVKESAPLSTRV--EKKESLPIIDVDE 249  
Qy 430 FALITPGDPRIVYISSGCATRLFEALVGAVPVVLGEOVLQYQDMLQWNEALVVPKPR 489  
Db 250 IE-----BGLVVRHREHGIAIEG--IVRLKGVILREYVL---KLKEDALLVPIERK 297  
Qy 490 VTEVHFLRLSLSDSDLLAMRQGRFLM-ETVYSTADSIENFVLAIRTRIQIPAPPIRE 548  
Db 298 IDVHV---KYGDPQSVQKLDKMRNGKWKOTLKKVRDIEKKIKELVELYMK-----RQ 348  
Qy 549 AAALPIHRSKGAAGTDPNMANNGDLDLGPVETEPYPASRYLRNFTLYTDFPRSNWCAP 608

Db 349 A-----QGLSLRGPDELEK-----AESFYIETPOQOSIEEVLSD----- 386  
Qy 609 GPHLPHTPPDPVLPSEANFLGSGTGRPIGAGSGSGKEFOALGANNPREQFTVVM 668  
Db 387 -----LASEKPMDRLLCGDA-----GVKTEVALRAAFVAVSGQVAVLVP 428  
Qy 669 T-----YE-----REVLNMS-----LELNLPPLNKKVYVWVNSPKL 701  
Db 429 TTVLAKHYENFKERMEPPGVKVELLDSRTAREKKEIETKL-----KGEIDIIITHTSL 484  
Qy 702 PSEDLMPDGVPI 716  
Db 485 LNERIEFSDGLVIT 499  
RESULT 14  
150618  
C:fps proto oncogene - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 18-Jun-1999  
C:Accession: 150618  
R: Huang, C.C.; Hammond, C.; Bishop, J.M.  
J. Mol. Biol. 181, 175-186, 1985  
A:Title: Nucleotide sequence and topography of chicken c-fps. Genesis of a retroviral on  
A:Reference number: 150618; MUID:85100838  
A:Accession: 150618  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-824 <HUG>  
A:Cross-references: EMBL:X02266; NID:g63203; PIDN:CA26155.1; PID:g871043  
C:Genetics:  
A:Gene: c-fps  
A:Initons: 75/3; 133/3; 166/1; 227/2; 273/2; 313/2; 354/2; 416/3; 442/3; 512/3; 553/3; 5  
C:Superfamily: protein-tyrosine kinase fps; protein kinase homology; SH2 homology  
C:Keywords: ATP  
F:462-547/Domain: SH2 homology <SH2>  
F:561-823/Domain: protein kinase homology <kin>  
F:569-577/Region: protein kinase ATP-binding motif  
Query Match 2.5%; Score 119.5; DB 2; Length 824;  
Best Local Similarity 18.8%; Pred. No. 3.2;  
Matches 162; Conservative 114; Mismatches 291; Indels 293; Gaps 41;  
Qy 79 VLDCIRRESVSEELDL-----EAKROELNSETAKLNLKTEACKSTENAKODLLQ 130  
Db 110 IRDQOLRKAFSEQMOOLSOEYARTTOQEMEKLAQYRSLVRDSTQAKRKQOASKDKER 169  
131 LKNVISTEHSYKELMANONPKLSPIRLPEKDDAGLPPKATRGCRNLNCPYSKCP 190  
Db 170 EKAKERYSLMKLYALHNDVYL-----AVRAAALHHHHHYORALP 210  
Qy 191 TSGPVVYVYSDQF-----VFGSY--LDPLVKAQFQATARANVVTENADACLYILV- 242  
Db 211 TLHSLSYSLQOEMVLYLKEILGECSISSLVQEDVLAIHQEVNAHVEIDPATYSSFVQ 270  
Qy 243 -----GEMQEPVLLPAEL-----EKQDYSLEPMFTGDHNNHYIMLSKRS 283  
Db 271 CHRYDSEVPRAVTPDESILLETESLEPGLQNLLETYSVQHSLSJTEELLASREAVSS 330  
Qy 284 TQNLIVNVSICRAMVAVOSTFTTVQYRGRFDLVSPVYHNSSEPMELPROVPVKRYTLF 343  
Db 331 KEQVWMLQ-----VELR-GEELALSP----- 351  
Qy 344 TFOGKIESTL--RSSLOARFSEEMEGDPPADYDRITLAKAVD---SKLDQVLEF 398  
Db 352 ---GERVHLGKRGLOEA---QOOLGVLCAQ-----AKLOQRMLANKLAEILOSE- 398  
Qy 399 TCKNQPRPSLPTF---MALGGERDRLELLKLSFALLITPGDRVLISSGCATRLFEAL 455  
Db 399 ---EPPALPLQDRDSQVCSVDERSGVTALETIKNHI-----SCIFSPRSL- 443  
Qy 456 EVGAVPVVVLGEOVLPYODMLQNHQALVLPKPRVTEVYHPLRLSLSDLLAMRQGR-- 513

Db 444 ---PPVPLIPEVOKPLCQO-AMTHGAI-----PR-SEVQELK--CGSDGLVRESQKOE 492  
Qy 514 -----FLME-----TYFSTADSI-----FNTVLAIRTRIQIPAPRIREAAAEIPIHRS 557  
Db 493 YVLSVMDGQPRHFIIOAONLVRLEBDGFTPIPLDHLQ-----SQQPIRKRS 543  
Qy 558 G---KAAGTDPNMANGDLDG-----PV-----ETEPYAS 586  
Db 544 GIYVLTAVLKDKNVLIHEDVLGERTIGRNGFEVSGRLRADNTPVAVKSCRETLPELK 603  
Qy 587 PRVLRNFTLVTPDFYRSMNCAPGFHLFPHPPPVLPSEAKFLSGTGRPI----- 639  
Db 604 AKFLQEARL-----LKQYN-----H-----PNIVRLGVCTQKQPIYVWELV 641  
Qy 640 -GGAGSGSGKEFOALGANNPREQFTVVMVLYEREVLMNSLERLNLPLNKKVWVNS 698  
Db 642 QCG-----DFLSFLRSEGRH-----LKKELIKMENNAAQMEYLE----- 677  
Qy 699 PKLPSEDLMPDGVPIVYVYVTEKNSLNRPVWNEIETAILSIDDAHLRHDEIMGF 758  
Db 678 ---SKHCIRHDLAARNCLV-TEKNT-----LKISD-----F 704  
Qy 759 RVMPREARDRIYVGFGRYH-----AMDIPHQSMVNSNYSCELSVLTGAFFHKYY--AYL 812  
Db 705 GMSRQEDGVYASTGKAKQIPVKWTAPBA--LNYGRVSSSDVWSFGILLMEAFSLGAVP 762  
Qy 813 YSYVMPQAIMDWDVETINC 832  
Db 763 YANLSNQOTREAIQGVRL 782  
RESULT 15  
545129  
VPS27 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein N2038; protein YNR006w  
C:Species: Saccharomyces cerevisiae  
C>Date: 03-May-1994 #sequence\_revision 02-Aug-1994 #text\_change 23-Mar-2001  
C:Accession: S45129; S48343; A57274; S59736; S63332  
R:Verhasselt, P.; Aert, R.; Voelt, M.; Volckaert, G.  
Submitted to the EMBL Data Library, January 1994  
A:Description: Twelve open reading frames revealed on the 23.6 kbp segment flanking t  
A:Reference number: S45119  
A:Accession: S45129  
A:Molecule type: DNA  
A:Residues: 1-622 <VER>  
A:Cross-references: EMBL:X77395; NID:g496717; PIDN:CA545474.1; PID:g496723  
R:Verhasselt, P.; Aert, R.; Voelt, M.; Volckaert, G.  
Yeast 10, 1355-1361, 1994  
A:Title: Twelve open reading frames revealed in the 23.6 kb segment flanking the cent  
A:Reference number: S48338; MUID:95208356  
A:Accession: S48343  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-622 <VE2>  
A:Cross-references: EMBL:X77395; NID:g496717; PIDN:CA545474.1; PID:g496723  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1994  
R: Piper, R.C.; Cooper, A.A.; Yang, H.; Stevens, T.H.  
J. Cell Biol. 131, 603-617, 1995  
A:Title: VPS27 controls vacuolar and endocytic traffic through a prevacuolar compartm  
A:Reference number: A57274; MUID:96042307  
A:Accession: A57274  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-320, 'NV', 323-622 <PIP>  
A:Cross-references: GB:U24218; NID:g785066; PIDN:AA96002.1; PID:g785067  
R: Piper, R.C.; Cooper, A.A.; Yang, H.; Stevens, T.H.  
Submitted to the EMBL Data Library, April 1995  
A:Reference number: S59736  
A:Accession: S59736  
A:Molecule type: DNA  
A:Residues: 1-320, 'NV', 323-622 <PIW>  
A:Cross-references: EMBL:U24218; NID:g785066; PID:g785067



R:Aert, R.; Verhasselt, P.; Voelt, M.; Volckaert, G.  
submitted to the Protein Sequence database, April 1996  
A:Reference number: S62910  
A:Accession: S63332  
A:Molecule type: DNA  
A:Residues: 1-622 <AER>  
A:Cross-references: EMBL:Z71620; NID:q1302476; PID:e239781; PID:q1302478; MIPS:YNR006w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:VPS27; VPT27; VPL23; SSV17  
A:Cross-references: SGD:S0005289; MIPS:YNR006w  
A:Map position: 14R  
C:Function:  
A:Description: controls vacuolar and endocytic traffic through a prevacuolar compartment  
C:Keywords: membrane trafficking

Query Match 2.4%; Score 118.5; DB 2; Length 622;  
Best Local Similarity 20.0%; Pred.No.2.4;  
Matches 85; Conservative 68; Mismatches 148; Indels 125; Gaps 15;

```
QY 15 GGQTCMLKMSNRIRLTWLTFTLVLFEPPL-----IAHYLTTLDEADEAGKR----- 63
Db 156 GGVFCEHSSNSIPLPLDLG-----IYEPYRVCDSCEFDYDLKRHDSSKSKKRRHKK 248
QY 64 ----IFGPRVGNELCEVKNHVIDLCRIRESVPELL-----QLEAKROELNSETAKLNK 113
Db 249 KDRDYSTEDEEELIRKAIETSLKESRNSASSEPIVPYVESKNEVKROIEEE-EDPDLK 307
QY 114 IEACKRSIENAKODLLQLNKVNISQTEHSYKEIQAQN---OPKLSLPIRLPEKDDAGLPP 170
Db 308 -AAIGESLREAEAKLR-----SERQKASROMQOPQSPQPIHSDVLSDEKDS----- 357
QY 171 PKATRGCLHNCFDYSCPLTSGFPVYVYDS-----DQFVFGSYLDPLVYKQAFQA 220
Db 358 -----IYMFASLVEKMKSRPLNEILEDKLNLAQRVAFAS 392
QY 221 TARANVYVENADIACLYIIVGEMQEPVLRPALEKQL-----YSLPHMRTDGHN 272
Db 393 KARLVYALNDKAAQYKNTLIENNGKISLETIMNTYDRLLQOLQISINLSQOYTLPOVPSDPYN 452
QY 273 HVIINLSRKSPTONL--LYNVSTGRAMVAOSTFYTVYRPGFDLVSPLVHAMSEPNFME 330
Db 453 YLTENVQNPAPESYQTPPLQQLSSH-----QYKPOQDVSRQOSVAKANSPT--- 497
QY 331 IPPQVPVKKYLTFTFGGKIESLRS-----SLQEARSFEEEMEGDDPPADYDRIIATLK 384
Db 498 -----TNIDHLKTIQVTPHAQCKPQSHVELAPSDPPYPRKEAEDEGDTQ 540
QY 385 AVQDSK 390
Db 541 AVQDEE 546
```

Search completed: September 4, 2002, 14:59:02  
Job time: 121 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:00 : Search time 18.3 Seconds  
(without alignments)  
1226.619 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873

Sequence: 1 MTGYTMLRNGAGNGGQTCM.....DSVLEKTRLPDKTKCFKFI 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
231628 seqs, 24425594 residues

Database: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	19.3	728	4	US-08-915-337-2
2	115.5	2.4	3075	2	US-08-460-309-5
3	115.5	2.4	3075	2	US-08-125-077-5
4	109.5	2.2	1572	2	US-08-290-731C-5
5	109.5	2.2	1596	3	US-09-356-952-3
6	108	2.2	873	3	US-09-187-331-6
7	108	2.2	873	4	US-09-470-946-6
8	108	2.2	925	2	US-08-392-946-1
9	108	2.2	925	2	US-08-504-169-1
10	108	2.2	925	5	PCT-US94-14893-1
11	103	2.1	2327	6	5455158-1
12	102.5	2.1	829	1	US-07-670-611-2
13	102.5	2.1	829	1	US-08-220-674-2
14	102.5	2.1	829	1	US-08-445-186-2
15	102.5	2.1	829	1	US-08-446-549-2
16	102.5	2.1	829	2	US-08-446-550-2
17	102.5	2.1	1093	4	US-09-315-793-52
18	101.5	2.1	3248	1	US-08-328-254-6
19	101.5	2.1	3248	1	US-08-353-700-1
20	101.5	2.1	3248	5	PCT-US95-16216-1
21	100.5	2.1	800	2	US-08-469-537A-72
22	100.5	2.1	800	2	US-08-469-537A-78
23	100.5	2.1	937	2	US-08-469-537A-105
24	100	2.1	2512	2	US-08-801-263A-9
25	99.5	2.0	428	1	US-09-102-248-9
26	99.5	2.0	428	1	US-08-785-050-2
27	99.5	2.0	428	2	US-08-996-798-2

28	99.5	2.0	655	4	US-09-298-724-2	Sequence 2, Appl
29	99	2.0	1089	1	US-08-180-195-36	Sequence 36, Appl
30	99	2.0	1089	1	US-08-477-329-36	Sequence 36, Appl
31	99	2.0	1089	2	US-08-475-458-36	Sequence 36, Appl
32	99	2.0	1089	3	US-08-980-400-36	Sequence 36, Appl
33	99	2.0	1089	4	US-09-583-459A-36	Sequence 36, Appl
34	99	2.0	1089	4	US-09-583-210-36	Sequence 36, Appl
35	99	2.0	1089	4	US-09-583-449A-36	Sequence 36, Appl
36	99	2.0	1579	3	US-08-755-587-184	Sequence 184, App
37	99	2.0	2329	3	US-08-755-587-16	Sequence 16, App
38	99	2.0	2386	2	US-09-016-366A-12	Sequence 12, Appl
39	99	2.0	2446	2	US-08-551-356-2	Sequence 2, Appl
40	99	2.0	2446	5	PCT-US93-12687-2	Sequence 2, Appl
41	99	2.0	3418	2	US-08-639-501-2	Sequence 2, Appl
42	99	2.0	3418	2	US-09-044-946-2	Sequence 2, Appl
43	99	2.0	3418	3	US-09-044-908-2	Sequence 2, Appl
44	98.5	2.0	1110	1	US-08-118-441-29	Sequence 29, Appl
45	98.5	2.0	1110	3	US-08-338-579A-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
US-08-915-337-2  
; Sequence 2, Application US/08915337  
; Patent No. 6287802  
; GENERAL INFORMATION:  
; APPLICANT: Deng, Fan & Xia  
; TITLE OF INVENTION: EXT2 Gene  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/915,337  
; FILING DATE: August 21, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: China - 96121928.9  
; FILING DATE: October 21, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: William T. Han  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50019  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 270-5219  
; TELEFAX: (610) 270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 728 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; US-08-915-337-2

Query Match 19.3% Score 939. DB 4: Length 728.

Best Local Similarity 31.6%; Pred. No. 1.4e-86;  
Matches 257; Conservative 126; Mismatches 256;  
Indels 174; Gaps 30.

[illegible]

RESULT 2  
 US-08-460-309-5  
 ; Sequence 5, Application US/08460309  
 ; Patent No. 5837496  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Engvall, Eva  
 ; APPLICANT: Leivo, Ilmo  
 ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin  
 ; TITLE OF INVENTION: Fragments and Uses thereof  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell and Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/460,309  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/125,077  
 FILING DATE: 22-SEP-1993  
 APPLICATION NUMBER: US PCT/US 94/10730  
 FILING DATE: 21-SEP-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/472,319  
 FILING DATE: 30-JAN-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/919,951  
 FILING DATE: 27-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-LA 9721  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3075 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-460-309-5

Query Match	2.4%	Score 115.5	DB 2	Length 3075
Best Local Similarity	20.1%	Pred. NO. 0.27		
Matches 120	Conservative 102	Mismatches 237	Indels 139	Gaps 27

[illegible]

Db 2103 KLISQ--ARKQASIVA---VSADRCIRAYOPQISSTNTNTLTINVTQEPDNLFF 2157

Qy 497 LRLSDSDLLAMR-ROGR--FLMETYFSTADSIPTVLAMIRRIQIIPAAREEAAEI 553

Db 2158 LGSSTASDFLAVEMRRCRVAFVFLMD-----LGSSTRLEPDPDPIDNRHMSI 2204

Qy 554 -----PHRSGKAAGTDBNMADNDL-----GVEYTERP 582

Db 2205 HVARFGNIGLSYKEMSSNOKSPTKTSKSPGT-ANVLVDVNNSTLMFVGLGQIKKSP 2261

RESULT 3

US-08-125-077-5

Sequence 5, Application US/08125077

Patent No. 5872231

Patent No. 5872231 5840863

GENERAL INFORMATION:

APPLICANT: Engvall, Eva

APPLICANT: Lelivo, Ilmo

TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/125,077

FILING DATE: 22-SEP-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US 94/10730

FILING DATE: 21-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/472,319

FILING DATE: 30-JAN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/919,951

FILING DATE: 27-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LA 9721

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ. ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3075 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-125-077-5

Query Match 2.4%; Score 115.5; DB 2; Length 3075;

Best Local Similarity 20.1%; Pred. No. 0.27;

Matches 120; Conservative 102; Mismatches 237; Indels 139; Gaps 27;

Qy 71 NELCEVAVHD-LCRIRSESVSEELQLEAKRQELNSIAKLKIECKKSIENAKODL 129

Db 1717 NATLELKAADLLSQIOENQOKPEELEVLEKASHVLSKHNNELKAALVREAEMQ 1776

Qy 130 QAKNVI-----SOTREHYKELMAQONPKLSIPRIILPEKNDAGIPPKATRGCGLHNCFD 184

Db 1777 ESNHLLMVANNALREFSDKRLHVOEEONLTSELV---OGRLIDAAAOQTDAVADALE 1832

Qy 185 YSRCPLTSGFPVYVDSDFVGS-----YLDPVYQAFQATARANYVTEN--ADIACL 227

Db 1833 HLE-----DHQDKLLMSAKIRHHIDDLVMHMSQRNAVLDLYRAEDHATEFORL 1881

Qy 238 YVLVGMGEPPVULPRAELEKOLYSLPHWRTDG---HNHYIINLSKRSQNLNLY--V 291

Db 1882 ADVLYSGLEN---IRNVSLNATSAAYVHNYIQSLIEESEELARDAHRTVETSLSESLV 1938

Qy 292 STGRAMVAOSTFYTVOYR-----PGFDLVVSPLVHMS--EPNFMEIPQO----- 334

Db 1939 SNGKAIVORSRFLKEBNNLKRKLPGIALLELSELKRNKTNFOENAVETITQTESLILR 1998

Qy 335 -VP-----VKRRYLETFQGEKIESLRSSLOARSEFEEMEGDPPADYDRIATLK-- 384

Db 1999 AIEGIRDKAKTKRELAT---SASQSAVSTLRVAGLSQELMTSAS--LSRVNTTLRET 2053

Qy 385 --AVQDSKLDQVLEFRCN---QPKRSLPTBVALCGERDRLELKLSTFALLITPGDP 439

Db 2054 HOLLQDSTMATLLAGRKRVKDEIQAKYLF-----DRUKPLKMEENLSRNLSET 2102

Qy 440 RLVISSGATRLFEALEVGAVPVVLGBOVLPYQDML---QMNDEALVYPRKPRVTEVHFL 496

Db 2103 KLISQ--ARKQASIVA---VSADRCIRAYOPQISSTNTNTLTINVTQEPDNLFF 2157

Qy 497 LRLSDSDLLAMR-ROGR--FLMETYFSTADSIPTVLAMIRRIQIIPAAREEAAEI 553

Db 2158 LGSSTASDFLAVEMRRCRVAFVFLMD-----LGSSTRLEPDPDPIDNRHMSI 2204

Qy 554 -----PHRSGKAAGTDBNMADNDL-----GVEYTERP 582

Db 2205 HVARFGNIGLSYKEMSSNOKSPTKTSKSPGT-ANVLVDVNNSTLMFVGLGQIKKSP 2261

RESULT 4

US-08-290-731C-5

Sequence 5, Application US/08290731C

Patent No. 5843646

GENERAL INFORMATION:

APPLICANT: BOWTELL, David Douglas Lawrence

TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE

TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS

STREET: 2100 PENNSYLVANIA AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

ZIP: 20037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,731C

FILING DATE: 17-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU93/00068

FILING DATE: 17-FEB-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PL0921/92

FILING DATE: 17-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: KIT, Gordon

REGISTRATION NUMBER: 30,764

REFERENCE/DOCKET NUMBER: 0-36066

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060

TELEFAX: (202) 293-7860

TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1572 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-290-731C-5

Query Match 2.2%; Score 109.5; DB 2; Length 1572;  
 Best Local Similarity 19.6%; Pred. No. 0.33; Indels 313; Gaps 54;  
 Matches 198; Conservative 142; Mismatches 359;

QY 76 VKHVLDCRIRSEVSEELLQLEAKROELNSELAKNLKIEA--CKSIENAKQDLOLQKN 133  
 DB 77 LKVLQVHPRYTAKEDALLY-----VEKICLRLLMLCAKPLPHSVQDVEEKNV 126  
 QY 134 -----VISQTEHSYKELMAQONQPKLSLP--IRLLPEKDDAGLPPEKATRGCRHLNCPD 184  
 DB 127 KSPFAPIDQMALNKAKEVINSKRRKSVLPTEKVTLLQKD-----VLQ 169  
 QY 185 YSRCPPLSGFPVYVYSDQFVFGSYL---DPLVKQAFQATARANYVTENADIACLIV- 239  
 DB 170 YKIDSSVSAGFLVAVL---EYISADILKMACGVYIKIAHCETTKEDIEVVMADRVLMQDL 226  
 QY 240 -----ILVGMQEPVVLRAPELEKOLYSLPH---WRTDGHNVITINLSRK-----SD 283  
 DB 227 NQSEAHILPSPILSPAPORASATYEETVKELIHDEKQYORD--LHMIVFREELVKIYSD 284  
 QY 284 TQNL-----LYNVS-----TGRAM--VAQSTFYVQYRPG 311  
 DB 285 PRELEPFSNIMDIYEVTYVLLGLSLEDVIMSQEOSAPCVSGFEELAEAEFDYKKYA 344  
 QY 312 FDLYV---SPLVHAMSEPNMEIP-----PO---VPVKRYLFTFGGEKI 350  
 DB 345 YDVTYSQASRDALNNLSKPGASSLTAGHGRDAVKYLLPRLLVLPICHAFFYDYIKHL 404  
 QY 351 ESLSLSQEARSPFEEMEGDPPADYD--DRIATLKAQVDSKLD-----QVLVEFTCKNOP 404  
 DB 405 KDLSSSDDIIESFOVGLLPHLCHDEKVMASLSKERQVPSGRVRQOLAIERTRELQ 464  
 QY 405 KPSLPTFW-----ALCGE--REDRLLEL-----KLSTF--ALITPGDPRLVIS 444  
 DB 465 KVE---HMEKDVQGNQNEFIREDSLSKSGKRIWSEKRVFLFDGLMVLCAKNTKQTP 521  
 QY 445 SGCAT-----RLFEALEVGAAPV-----VLGEQVOLPY-----QDMLQWNE 480  
 DB 522 SAGATADYRLKEKVFMRKRVINDRPDSDDLKNSFELAPRMQPRIVLTAKNAQKHDMMA 581  
 QY 481 AALVYPRPRTVEHF--LNLSSDSDDLAMRROGRFLMEYI--ESTADISIFNTYLAMITR 537  
 DB 582 DLMVITKSMIDRLDSIIDDIERKHPLRMSP-----EYKFAVPSGQNIIVL----- 630  
 QY 538 IOIPADIRREAAEIPHRSGKAAGTDPNMAANDGLDLPVETEPPEAPSPRYLNNFTLV 597  
 DB 631 -----EBRESAGVIMIGATL-----CKLIERLYHIYADPFEVVFTEFLT- 670  
 QY 598 TDRYRSNCAAPGPHLEPHTPDPVLPSEAKFLGSGTGFRPIGGAGSGKEFOALG- 656  
 DB 671 ---YR-YFCSPQOOLLQLLVERFNIPDS-LYQDGTG-----AGAGGMC-----GVGCD 714  
 QY 657 ---NPRE-----QFTYV-----MLTYERE-----EVLMSLEKLNG 685  
 DB 715 KEKHNHREDMKRYRREYQVQFRVNLVLRHVDHHPYFEKDPMLLEKLLNFEHLVNG 774  
 QY 686 LPLINKVVVVNSPKLSEDLMPDGVPLMVYVTEKNSLNNRFL---PNNLEIETEILS 742  
 DB 775 -KSMRK-----WVDSYLVKIQKKNDEKSNKKTIVAYVGHDPPEIHHLS 817  
 QY 743 IDDD-----AHLRHDEIMFGFRVWRAR--DRIYGFGRYHAMDIPQSWL---YN 788  
 DB 818 VPNDDELTLTLHLPLLELAQUTLLEFEYKKNVKSSELVGSWTKKDKVEKSPNLKIKMKT 877

QY 789 SNVSCELSMVLGAAFFHKKYAYLYSYVPOAIRDMVDEYINCEDIAMNFLTSHTRKP 848  
 DB 878 TWNTWIEKSTIEAEYERELA-----IMQRAIEVMM---VLEELNNNGILSIYAANGT 929  
 QY 849 IKV-TSRWTFRCGCPQ-----ALSHDD-SHFERNK-----CINFP 883  
 DB 930 ASVYRLRMTFQ--GLPERYRKFLRECRLELSDDLKKYQERLRSINPCVPFF 979

RESULT 5  
 US-09-356-952-3  
 ; Sequence 3, Application US/09356952  
 ; Patent No. 6117663  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boriack-Stodin, Ann  
 ; APPLICANT: Margalit, S. M.  
 ; APPLICANT: Bor-Sogli, Dafna  
 ; APPLICANT: Cole, Philip  
 ; APPLICANT: Kuriyan, John  
 ; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
 ; TITLE OF INVENTION: THEREOF  
 ; FILING DATE: 1998-07-19  
 ; CURRENT FILING DATE: 1999-07-19  
 ; EARLIER FILING DATE: 1998-07-19  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: patentin Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 1596  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-356-952-3

Query Match 2.2%; Score 109.5; DB 3; Length 1596;  
 Best Local Similarity 19.6%; Pred. No. 0.34;  
 Matches 198; Conservative 142; Mismatches 359; Indels 313; Gaps 54;

QY 76 VKHVLDCRIRSEVSEELLQLEAKROELNSELAKNLKIEA--CKSIENAKQDLOLQKN 133  
 DB 77 LKVLQVHPRYTAKEDALLY-----VEKICLRLLMLCAKPLPHSVQDVEEKNV 126  
 QY 134 -----VISQTEHSYKELMAQONQPKLSLP--IRLLPEKDDAGLPPEKATRGCRHLNCPD 184  
 DB 127 KSPFAPIDQMALNKAKEVINSKRRKSVLPTEKVTLLQKD-----VLQ 169  
 QY 185 YSRCPPLSGFPVYVYSDQFVFGSYL---DPLVKQAFQATARANYVTENADIACLIV- 239  
 DB 170 YKIDSSVSAGFLVAVL---EYISADILKMACGVYIKIAHCETTKEDIEVVMADRVLMQDL 226  
 QY 240 -----ILVGMQEPVVLRAPELEKOLYSLPH---WRTDGHNVITINLSRK-----SD 283  
 DB 227 NQSEAHILPSPILSPAPORASATYEETVKELIHDEKQYORD--LHMIVFREELVKIYSD 284  
 QY 284 TQNL-----LYNVS-----TGRAM--VAQSTFYVQYRPG 311  
 DB 285 PRELEPFSNIMDIYEVTYVLLGLSLEDVIMSQEOSAPCVSGFEELAEAEFDYKKYA 344  
 QY 312 FDLYV---SPLVHAMSEPNMEIP-----PO---VPVKRYLFTFGGEKI 350  
 DB 345 YDVTYSQASRDALNNLSKPGASSLTAGHGRDAVKYLLPRLLVLPICHAFFYDYIKHL 404  
 QY 351 ESLSLSQEARSPFEEMEGDPPADYD--DRIATLKAQVDSKLD-----QVLVEFTCKNOP 404  
 DB 405 KDLSSSDDIIESFOVGLLPHLCHDEKVMASLSKERQVPSGRVRQOLAIERTRELQ 464  
 QY 405 KPSLPTFW-----ALCGE--REDRLLEL-----KLSTF--ALITPGDPRLVIS 444  
 DB 465 KVE---HMEKDVQGNQNEFIREDSLSKSGKRIWSEKRVFLFDGLMVLCAKNTKQTP 521  
 QY 445 SGCAT-----RLFEALEVGAAPV-----VLGEQVOLPY-----QDMLQWNE 480

```

Db 522 SAGTATADYDLKEKYEFRARRDINDRPSDDLKNSFELAPMPRIYVLTAKNAQKHMMMA 581
OY 481 AALVPRKRVTEVHF--LRLSLSOSDILAMRGRFLMETY-ESTADSIENVYLAIRTR 537
Db 582 DLMVITKSMILDRILDSILQDIERKHPLRMSP-----EIKFAVPSGDNIVL----- 630
OY 538 IQIPAAPIREMAAEIHRSGKAAGTDPNMADNDLDLGVEPEPPYASRYLRLNFTLV 597
Db 631 -----EERESAGVMIKATL-----CKLIERLYHIADPTEVTFLLT- 670
OY 598 TDFRSWNCAPRPHLPHTPDPVLPSEAKFLGSGTFPRIGGAGSGKEFOALG- 656
Db 671 ---YR-YFCSPQDLQLLVERFNIIPDS-LYQDTGT-----AGAGCG- 714
OY 657 ---NPRE-----OFTVY-----MLTYERE-----EYLMNSLERLNG 685
Db 715 KEHNSHREDMKRYKREYQPVGFRVLNVLRLHWYDHFDEKDPMLLEKLNLEHNG 774
OY 686 LPIYLNKVVVWNSPKLPSEDLMPDVGPIMVVTEKNSLNRL---PNNEITEALIS 742
Db 775 -KSMRK-----WVDSVLKIVORKNEQEKSNKIVYAYGHPPEIEHLS 817
OY 743 IDDD-----AHLRHEIMGFRVYRREAR-DRIVGFGRYHAMDIPQSM- 788
Db 818 VPNDITLLTLHLELAKOLTLLEFEMTKVVKPSSELVGSFMTKKDKVEKSPNLKIKHT 877
OY 789 SNSCELSNVLTCGAFFHKYAYLYSVMPQAIRDNDVDEXINCEDIAMNPLVSHITRPP 848
Db 878 TWTWRIEKSITEAENEERLA-----IMQRAIEVMM---VMLELNFNGILSIYAAMGT 929
OY 849 IKV-TSMWTRCQCGPO-----ALSHDD-SHFERHK-----CINEF 883
Db 930 ASYVRLMWTQ--GLPERYKRFLECRELSDDLHKYQERLRSINPCVBF 979

```

```

RESULT 6
US-09-187-331-6
; Sequence 6, Application US/09187331
; Patent No. 6043056
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/187,331
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 873
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g189650
US-09-187-331-6

```

```

Query Match 2.28; Score 108; DB 3; Length 873;
Best Local Similarity 23.8%; Pred. No. 0.17;
Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;
OY 541 PAAPIREMAAEIHR--SGKAAGTDPNMADNDGLDGPVETEPYASRYLRLNFTLV 598
Db 401 PAARLR--PSDVPDKYSTRYEGIANRLSCR-----EPNQHFKPYLKLFLPKRL 447
OY 599 DFRSNWNCAPRPHLPHTPDPVLPSEAKFLGSGTFPRIGGAGSGKEFOALGNV 658
Db 448 HFAKSDRIEPLTFYLDQWQL-ALNPSEKRYCGSGF-----HSDNVVFSMQALFVGYG 500

```

```

OY 659 PROQTVVMLTYEREY--LMNSLERL-----NG-----LPIYLNKVVVWNSPKLPSEDL 707
Db 501 PGFKHGIEADTFENIEYVNLMDLNLTPAPNNGTSGLSLHNLKNPY--TPKHPE--V 536
OY 708 WPDIGVIMVVRTEKNSL-----NNRFLPMNEIEIEALISIDDAHLRHEIMGFRVWRE 763
Db 557 HPLVQCFP--TRNPRDLGSCNPSILPIEDPQOFNLYAEEKIIKHETLPY- 608
OY 764 ARDRIVGFGRYHAMDIPHOSWLYNSVSCELSNVLTCGAFFHKYVA-YLYSVMPQAIR 822
Db 609 -RRRV-----LQKENTICLLSQ-----HQFMSGYQDILMLPMTS 642
OY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKVTSMWTRCQCGQALSHDSSHFERHKCI 880
Db 643 YTVDRNDSFSTEDFS-NCLYQ-----DERIPLSPV-----HKC- 674
OY 881 NFF---YKV-YGYM 890
Db 675 SFYKNNTKVSGYGL 688

```

```

RESULT 7
US-09-470-946-6
; Sequence 6, Application US/09470946
; Patent No. 6358923
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/470,946
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: US 09/187,331
; EARLIER FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 873
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g189650
US-09-470-946-6

```

```

Query Match 2.28; Score 108; DB 4; Length 873;
Best Local Similarity 23.8%; Pred. No. 0.17;
Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;
OY 541 PAAPIREMAAEIHR--SGKAAGTDPNMADNDGLDGPVETEPYASRYLRLNFTLV 598
Db 401 PAARLR--PSDVPDKYSTRYEGIANRLSCR-----EPNQHFKPYLKLFLPKRL 447
OY 599 DFRSNWNCAPRPHLPHTPDPVLPSEAKFLGSGTFPRIGGAGSGKEFOALGNV 658
Db 448 HFAKSDRIEPLTFYLDQWQL-ALNPSEKRYCGSGF-----HSDNVVFSMQALFVGYG 500
OY 659 PROQTVVMLTYEREY--LMNSLERL-----NG-----LPIYLNKVVVWNSPKLPSEDL 707
Db 501 PGFKHGIEADTFENIEYVNLMDLNLTPAPNNGTSGLSLHNLKNPY--TPKHPE--V 536
OY 708 WPDIGVIMVVRTEKNSL-----NNRFLPMNEIEIEALISIDDAHLRHEIMGFRVWRE 763
Db 557 HPLVQCFP--TRNPRDLGSCNPSILPIEDPQOFNLYAEEKIIKHETLPY- 608
OY 764 ARDRIVGFGRYHAMDIPHOSWLYNSVSCELSNVLTCGAFFHKYVA-YLYSVMPQAIR 822
Db 609 -RRRV-----LQKENTICLLSQ-----HQFMSGYQDILMLPMTS 642
OY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKVTSMWTRCQCGQALSHDSSHFERHKCI 880

```

Db 643 YVVDNDSTEDFS-NCILYQ-----DERIPLSPV-----HRC- 674  
QY 881 NFE---VKV-YGYM 890  
Db 675 SFYKNNTKVSYGFL 688

## RESULT 8

US-08-392-946-1  
; Sequence 1, Application US/08392946  
; Patent No. 5939269  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: The Regents of the University of California  
; APPLICANT: Goldfine, Ira D.  
; APPLICANT: Grupe, Andrew  
; APPLICANT: Maddox, Betty A.  
; APPLICANT: Spencer, Steven  
; APPLICANT: Stewart, Timothy A.  
; TITLE OF INVENTION: Antagonists to Insulin Receptor Tyrosine  
; TITLE OF INVENTION: Kinase Inhibitor  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/392,946  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/182241  
; FILING DATE: 14-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kudinec, Jeffrey S.  
; REGISTRATION NUMBER: 36,575  
; REFERENCE/DOCKET NUMBER: P0875P1PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-8228  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 925 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
; US-08-392-946-1

Query Match 2.2%; Score 108; DB 2; Length 925;

Best Local Similarity 23.8%; Pred. No. 0.18;

Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

QY 541 PAAPIREAAAEIPIHR--SGKAAGTDPNMADNGLDGLGVPVTEPPYASPRYLRFNFTLTVT 598  
Db 453 PAAPLR--PSDVPDKYYSFNYEGIAIRNLSCR-----EPNQHFPPYLKHFLPKRL 499  
QY 599 DFRSNNACGAPFHLFHTFEDPVLPSSEAKFLGSGTGFRPIGGAGGSGKEFOALGCV 658  
Db 500 HFAKSDIEPLTLFDQOML-ALNPSEKRYCGSGF-----HESDVFSSNMQLFVGYG 552  
QY 659 PRQFTVVMLYEREVY--LANSLERL-----NG---LPYLNKVVVWMSPKLPSEDL 707  
Db 553 PGRKHGIEADTFENIEVYNNLMCDLNLTPAPNNGTHOSLHLLKNPY--TPKHPRK--V 608

QY 708 WPDIGVPIVNVTEKNSL-----NNRFLPWNIEIETAIISIDDAHLRHDEIMFGRWRE 763  
Db 609 HPLVOCPE--TRNPRDLGCSCNPISILPIEDFOTQFNLTVAEREKIHKHTLPYG----- 660  
QY 764 ARDRIVGFGRYHANDIPHQSLYNSNYSCELSMWLTGAFFHKYYA--LYSYVVPQAIR 822  
Db 661 RPRV-----LQKENTICLSQ-----HOFMSGVSQDIIMPWTMS 694  
QY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKYSRMTFCPCGCPQALSHDHSFHERKCI 880  
Db 695 YVVDNDSTEDFS-NCILYQ-----DFRIPLSPV-----HRC- 726  
QY 881 NFE---VKV-YGYM 890  
Db 727 SFYKNNTKVSYGFL 740

## RESULT 9

US-08-504-169-1  
; Sequence 1, Application US/08504169  
; Patent No. 5968508  
; GENERAL INFORMATION:  
; APPLICANT: Goldfine, Ira  
; APPLICANT: Grupe, Andrew  
; APPLICANT: Henzel, William  
; APPLICANT: Maddox, Betty  
; APPLICANT: Spencer, Steven  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Harrison, Denise M.  
; TITLE OF INVENTION: Antibodies to Insulin Receptor Tyrosine Kinase Activation I  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/504,169  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/14893  
; FILING DATE: 28-Dec-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/182241  
; FILING DATE: 14-Jan-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kudinec, Jeffrey S.  
; REGISTRATION NUMBER: 36,575  
; REFERENCE/DOCKET NUMBER: P0875P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8228  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 925 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-504-169-1

Query Match 2.2%; Score 108; DB 2; Length 925;

Best Local Similarity 23.8%; Pred. No. 0.18;

Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

QY 541 PAAPIREAAAEIPHR--SGKAAGTDPNMADNGDLDGPEVETEPYASPRYLNFLLTVT 598  
 DB 453 PAARLR--PSDVPDKYSENYEGIANRLSCR-----EPNQHFXYLKHFLPKRL 499  
 QY 599 DFRSNWNCAGPRLPHPTFPDVLPSSEAKFLGSGTFRPGAGGSGKEFOALAGNV 658  
 DB 500 HFASQDRIEPLTFLDPOMOL-ALNPSEKRYCGSGF-----HGSDFVFSNMOLALFYGYG 552  
 QY 659 PROFTVVLTYREEV--LMSLERL-----NG----LPYLKVVVVMNSPKLPSEDL 707  
 DB 553 PGRHGIADTFENIEYNYLMCDLNLTPAPNNGTHOSLHLLKNPY--TPKHKE--V 608  
 QY 708 WPDIGPIMVVRTEKNSL---NNRFLPMNEIETALISIDDAHLRHDEIMGFYRWRE 763  
 DB 609 HPLVQCF--TRNPDMLGSCNPSILPIEDFQFNLTVAEKIKIHEFLPYG----- 660  
 QY 764 ARDRIVGPGRYHAMDIPOHSMWLYNSYSCELSVLGAAPFKHYA-YLYSVYMPQAIR 822  
 DB 661 -RRRV-----LQKENTICLSO-----HOFMSGYSDILMLPMTS 694  
 QY 823 DMYD--EYINCEDIAMNFLVSHITRKPRPIKVTSMWFRCPGCPQALSHDSSHFERHKKI 880  
 DB 695 YTVDRNDSTEDFS-NCLYQ-----DRIPLSPV-----HKC- 726  
 QY 881 NFF---VKV-YGYM 890  
 DB 727 SFYKNNTKVSYGFL 740

RESULT 10

PCT-US94-14893-1  
 ; Sequence 1, Application PC/TUS9414893  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genentech, Inc.  
 ; APPLICANT: The Regents of the University of California  
 ; APPLICANT: Goldfine, Ira D.  
 ; APPLICANT: Grupe, Andrew  
 ; APPLICANT: Maddux, Betty A.  
 ; APPLICANT: Spencer, Steven  
 ; APPLICANT: Stewart, Timothy A.  
 ; TITLE OF INVENTION: Antagonists to Insulin Receptor Tyrosine Kinase Inhibitor  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: palin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/14893  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/182241  
 ; FILING DATE: 14-JAN-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kudinec, Jeffrey S.  
 ; REGISTRATION NUMBER: 36,575  
 ; REFERENCE/DOCKET NUMBER: 875P1PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE:  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 925 amino acids  
 ; TYPE: amino acid

; TOPOLOGY: linear  
 PCT-US94-14893-1

Query Match 2.2%; Score 108; DB 5; Length 925;  
 Best Local Similarity 23.8%; Pred. No. 0.18;  
 Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

QY 541 PAAPIREAAAEIPHR--SGKAAGTDPNMADNGDLDGPEVETEPYASPRYLNFLLTVT 598  
 DB 453 PAARLR--PSDVPDKYSENYEGIANRLSCR-----EPNQHFXYLKHFLPKRL 499  
 QY 599 DFRSNWNCAGPRLPHPTFPDVLPSSEAKFLGSGTFRPGAGGSGKEFOALAGNV 658  
 DB 500 HFASQDRIEPLTFLDPOMOL-ALNPSEKRYCGSGF-----HGSDFVFSNMOLALFYGYG 552  
 QY 659 PROFTVVLTYREEV--LMSLERL-----NG----LPYLKVVVVMNSPKLPSEDL 707  
 DB 553 PGRHGIADTFENIEYNYLMCDLNLTPAPNNGTHOSLHLLKNPY--TPKHKE--V 608  
 QY 708 WPDIGPIMVVRTEKNSL---NNRFLPMNEIETALISIDDAHLRHDEIMGFYRWRE 763  
 DB 609 HPLVQCF--TRNPDMLGSCNPSILPIEDFQFNLTVAEKIKIHEFLPYG----- 660  
 QY 764 ARDRIVGPGRYHAMDIPOHSMWLYNSYSCELSVLGAAPFKHYA-YLYSVYMPQAIR 822  
 DB 661 -RRRV-----LQKENTICLSO-----HOFMSGYSDILMLPMTS 694  
 QY 823 DMYD--EYINCEDIAMNFLVSHITRKPRPIKVTSMWFRCPGCPQALSHDSSHFERHKKI 880  
 DB 695 YTVDRNDSTEDFS-NCLYQ-----DRIPLSPV-----HKC- 726  
 QY 881 NFF---VKV-YGYM 890  
 DB 727 SFYKNNTKVSYGFL 740

RESULT 11

545158-1  
 ; Patent No. 545158  
 ; APPLICANT: VOGEL, TIKVA; LEVANON, AVIGDOR; WERBER, MOSHE M.;  
 ; GUY, RACHEL; PANET, AMOS  
 ; TITLE OF INVENTION: FIBRIN BINDING DOMAIN POLYPEPTIDES AND  
 ; USES AND METHODS OF PRODUCING SAME  
 ; NUMBER OF SEQUENCES: 20  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/58,241  
 ; FILING DATE: 04-MAY-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 526,397  
 ; FILING DATE: 21-MAY-1990  
 ; APPLICATION NUMBER: 345,952  
 ; FILING DATE: 28-APR-1989  
 ; APPLICATION NUMBER: 291,951  
 ; SEQ ID NO:1:  
 ; LENGTH: 2327  
 545158-1

Query Match 2.1%; Score 103; DB 6; Length 2327;  
 Best Local Similarity 22.2%; Pred. No. 3.1;  
 Matches 119; Conservative 67; Mismatches 184; Indels 166; Gaps 32;

QY 257 EKOLYSLPHWRTD-----GHNHVIINSRKSQ--TONLIVNVSGRAMVA----- 299  
 DB 734 EPQYLDLPSTATSVNIPDLPGKRY-IVNYYQISEDEQSLLISTGTTAPDAPDPPTVD 792  
 QY 300 -OSTFYTVQY-RP-----GFDLVVSPLVHAMS-EPNFMETIPPOVPKR-----KYLTF 345  
 DB 793 QVDDTSIVVMSRQAVITGYRIYSPVSGSSSELNLPETANSVTLSDLOPGQVYNTI 852  
 QY 346 QG--EKIESLRSSLOEARSEEBEMEGDPADYDDRIATLAKAVODSKLDVYL-VEFTCKN 402



```

US-07-670-611-2

Query Match          2.1%; Score 102.5; DB 1; Length 829;
Best Local Similarity 29.4%; Pred. No. 0.55;
Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3.

QY 59 EAGRIFGPRVGNELCEVHYLDLCRIRESSEE--LLDLEAKROFINSEIAKLNKIIEAC 117
   | : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 EREDLDLEKLAKKACEESH---LMREHEDVOERTTLTKEEYITLHSLVIAELNKKIDRL 99

QY 118 KKSIEAKODLLQLNKVISQTEHSYKE---LMAQNQPKSLP 156
   : : : : | | : | | | | | | | | | | | | | | | | | | | | |
Db 100 QGTTIREDEKXSLRSELSQSHQHEVNEEDSRMDQDQTSVIP 141

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEERITELHVAELNKKIDRL 99  
QY 118 KSIENAKODLLQKLVISQTEHSYKE---LMAONQPKLSLP 156  
Db 100 QGTIREDEYSELRSLSOSQHVNEDSRMDQDQTSVSP 141

RESULT 14  
US-08-445-186-2  
; Sequence 2, Application US/08445186  
; Patent No. 5576422  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: White, Raymond  
; APPLICANT: Nakamura, Yusuke  
; TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of  
; TITLE OF INVENTION: Humans  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch et al.  
; STREET: 1001 G Street  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001-4597  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,186  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 530  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/220,674  
; FILING DATE: 31-MAR-1994  
; APPLICATION NUMBER: US 07/670,611  
; FILING DATE: 13-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.33981  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 829 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-445-186-2

Query Match 2.18; Score 102.5; DB 1; Length 829;  
Best Local Similarity 29.48; Pred. No. 0.55;  
Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;  
QY 59 EAGRIRGPRVGNELCEVKKHYLDLCRIRESYSE-LQLEAKKQELNSELAKLNKLTIEAC 117  
Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEERITELHVAELNKKIDRL 99  
QY 118 KSIENAKODLLQKLVISQTEHSYKE---LMAONQPKLSLP 156  
Db 100 QGTIREDEYSELRSLSOSQHVNEDSRMDQDQTSVSP 141

RESULT 15  
US-08-446-549-2  
; Sequence 2, Application US/08446549  
; Patent No. 5693536  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; APPLICANT: White, Raymond  
; APPLICANT: Nakamura, Yusuke  
; TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of  
; TITLE OF INVENTION: Humans  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch et al.  
; STREET: 1001 G Street  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001-4597  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,549  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/670,611  
; FILING DATE: 13-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.33981  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 829 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-446-549-2

Query Match 2.18; Score 102.5; DB 1; Length 829;  
Best Local Similarity 29.48; Pred. No. 0.55;  
Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;  
QY 59 EAGRIRGPRVGNELCEVKKHYLDLCRIRESYSE-LQLEAKKQELNSELAKLNKLTIEAC 117  
Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEERITELHVAELNKKIDRL 99  
QY 118 KSIENAKODLLQKLVISQTEHSYKE---LMAONQPKLSLP 156  
Db 100 QGTIREDEYSELRSLSOSQHVNEDSRMDQDQTSVSP 141

Search completed: September 4, 2002, 14:57:37  
Job time: 37 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:00 ; Search time 35.77 Seconds

(without alignments)  
2853.699 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873

Sequence: 1 MTGYTLRNGAGANGCGTCM.....DSVLEKRLPHDKTKCFKFI 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 747574 seqs, 111073796 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4873	100.0	919	20	AAV08481 Human exostosin-4
2	4873	100.0	919	21	AAV84669 Amino acid sequenc
3	4755	97.6	919	22	AAV84668 Rat Reg-binding pr
4	4720.5	96.9	918	21	AAV84668 Amino acid sequenc
5	2314	47.5	972	22	ABB70158 Drosophila melanog
6	1666.5	34.2	364	22	AAV84675 Rat Reg-binding pr
7	1436	29.5	262	21	AAV84675 Amino acid sequenc
8	1000	20.5	718	20	AAV28883 Human Ext-2 protei
9	998	20.5	663	21	AAV858946 Breast and ovarian
10	939	19.3	728	19	AAV44851 Hereditary multipl
11	871	17.9	746	20	AAV28862 Human Ext-1 protei

12	848.5	17.4	259	21	AAV84680 Amino acid sequenc
13	841	17.3	717	22	ABB61749 Drosophila melanog
14	796.5	16.3	760	20	AAV28884 Drosophila Tcy pro
15	795.5	16.3	760	22	ABB64881 Drosophila melanog
16	663.5	13.6	269	21	AAV84676 Amino acid sequenc
17	625	12.8	125	21	AAV84670 Amino acid sequenc
18	530	10.9	270	21	AAV84677 Amino acid sequenc
19	399.5	8.2	262	21	AAV84678 Amino acid sequenc
20	367	7.5	320	21	AAV82363 Human alpha-1,4-N
21	365	7.5	270	21	AAV84679 Amino acid sequenc
22	290	6.0	329	21	AAV84673 Arabidopsis thalia
23	199.5	4.1	123	21	AAV84673 Amino acid sequenc
24	197.5	4.1	124	21	AAV84672 Amino acid sequenc
25	180	3.7	121	21	AAV84674 Amino acid sequenc
26	176.5	3.6	120	21	AAV84671 Amino acid sequenc
27	160	3.3	103	20	AAV60148 Human endometrium
28	142	2.9	161	22	AAV01937 Human polypeptide
29	141	2.9	181	22	ABG30294 Novel human diagn
30	123	2.5	1378	22	ABG23678 Novel human diagn
31	123	2.5	1400	22	ABG09151 Novel human diagn
32	123	2.5	1788	22	ABG06749 Novel human diagn
33	123	2.5	1921	22	ABG06576 Novel human diagn
34	123	2.5	2563	22	ABG14767 Novel human diagn
35	122.5	2.5	537	21	AAV14498 Arabidopsis thalia
36	122.5	2.5	549	21	AAV14497 Arabidopsis thalia
37	121.5	2.5	415	21	AAV07301 Arabidopsis thalia
38	121.5	2.5	415	21	AAV46602 Arabidopsis thalia
39	121.5	2.5	423	21	AAV46602 Arabidopsis thalia
40	121.5	2.5	423	21	AAV41326 Arabidopsis thalia
41	121.5	2.5	755	22	ABG04875 Novel human diagn
42	120.5	2.5	2633	22	ABG06505 Novel human diagn
43	119.5	2.5	2901	22	ABG09763 Novel human diagn
44	119	2.4	613	22	ABG06752 Arabidopsis thalia
45	118.5	2.4	412	21	AAV26205 Arabidopsis thalia

## ALIGNMENTS

RESULT	1	ALIGNMENTS
AAV08481	ID	AAV08481 standard; Protein; 919 AA.
XX	AC	AAV08481:
XX	DT	29-JUL-1999 (first entry)
XX	DE	Human exostosin-4 protein.
XX	XX	Exostosin-4; human; cytosolic; immunosuppressive; anti-HIV; Ext-2;
KW	XX	treatment; cancer; tumour; genetic disorder; AIDS; diagnostic; screening;
KW	XX	acquired immune deficiency syndrome.
XX	OS	Homo sapiens.
XX	XX	WO925822-A1.
XX	PN	WO925822-A1.
XX	PD	27-MAY-1999.
XX	PF	14-NOV-1997; 97WO-CN00126.
XX	PR	14-NOV-1997; 97WO-CN00126.
XX	XX	(UYHU-) UNIV HUNAN MEDICAL.
PI	XX	Deng H, Fan C, Liu C, Ruan Q, Xia J, Xu L;
DR	XX	WPI: 1999-347470/29.
DR	XX	N-PSDB: AAV72378.
PT	XX	New Exostosin-4 gene useful in the treatment of cancer, tumors,
XX	XX	genetic disorders and AIDS

PS Claim 1; Page 24-27; 33pp; English.

CC This invention describes a novel human Exostosin-4 polypeptide which has  
 CC cytosolic, immunosuppressive and anti-HIV activity. Exostosin-4 has  
 CC homology with Ext-2 and is therefore expected to have similar biological  
 CC activities. Exostosin-4 and its modulators can be used for treatment of  
 CC cancer, tumours, genetic disorders and AIDS (Acquired Immune Deficiency  
 CC Syndrome), and in diagnostic/screening assays for such conditions.

XX Sequence 919 AA:

Query Match 100.0%; Score 4873; DB 20; Length 919;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGYTMLRNGAGNGGQCMLRMSNRIRLTWLSFTLVLFVLPPLIAHYLTTLDEADBA 60  
 DB 1 mtygmrlngagnggqctmlrwsnrirltwlsftlvlfvlppliahyltltldeada 60  
 QY 61 GKRIFGPRVGNELCEVKHVLDCRIRSVSEELLQLEAKROELNSEIAKLNLKIEACKS 120  
 DB 61 gkrifgprvgnelcevkhlvdlcrirsvseellqleakrgelnselaklnlkieacks 120  
 QY 121 IENAKQDLLQKLVISQTEHSHYKELMAONOPKLSPIRLLPEKDDAGLPPEKATGCRHL 180  
 DB 121 ienakqdllqklnvisqtehsykelmagqklsipirllpekddaglppekattrgcrhl 180  
 QY 181 NCFDYSRCPLTSGFPVYVYDSQVFGSYLDPVKAQAFQATARANVYTENADIACLYVI 240  
 DB 181 ncfdysrcpltsfgfpvyvdsqvfysyldplvkafqataranvytenadiaclyvi 240  
 QY 241 LVGEKQEPVLRPALELEKOLSLPHMRDGNHVNINLSRSDONLXNSTGAMVAQ 300  
 DB 241 lvgekepvrlrpalalekolslphmrtdgnhvninlsrdsdonlxnstgamaaq 300  
 QY 301 STFTYVOYRPGFDLVSPLVHAMSEPNFMEIPQVPRKYLFTFOGKISLSLSLQEA 360  
 DB 301 stfityvoyrpgfdlvsplvhamsepnfmeipqvpkrkylftfgekislslsqlea 360  
 QY 361 RSFEEMBGDPADYDRIITTLKAVODSKLDQVLEFTECKNOKPSLPTFMALGGERED 420  
 DB 361 rsfeembgdpadydriitltkavodskldqyleftecknokpslptfmalggered 420  
 QY 421 RLLELKSTFALITTPGPRLVISGGCATRLEFALGVAVVVLGEQVQLPYQMLQNE 480  
 DB 421 rllelkstfalittpgprlvissgcattrlefalgvavvvlgvqqlpyqmlqne 480  
 QY 481 AALVVPKRVTEVHFLRLSDSDLLAMRQGRFLMETESTADISFNTVLAIRTRIOI 540  
 DB 481 aalvvpkrvtevhflrlsdsdldlamrqrflmetestadsifntvlamirtrioi 540  
 QY 541 PAAPIREAAAEIRPHRSGKAAGTPPNADNGDLDLGEVETEPYPASPRYLKNTFLVTD 600  
 DB 541 paapireaaaeirphrsgkaagtppnadngdldlgevetepypasprylkntflvtdf 600  
 QY 601 YRSWNCAPGCFHLEPHRPPVPLPSEAKTSGSGRPPICGAGSGSKERQALGVNPR 660  
 DB 601 yrswncapgcfhlephrppvplpseaktsgsgrppicgagsgskerqalgvnpr 660  
 QY 661 EQFTVVMVLYERBEVLNLSLERLNGLPYLNKVVVVMNSPKLPSDDLMPDIGNVVRT 720  
 DB 661 eqftvvmvlyerbevlnlslerlnglpynlkvvvvnspklpsddlmpdignvrvrt 720  
 QY 721 EKNSLNRRFLPMEIETEAALISDDAHLRHDEIMFGFRVREARDRIVGFGRYHAMD 780  
 DB 721 eknslnrrflpmeieteaalisddahlrhdeimfgfrvreardrvfgfryhawn 780  
 QY 781 PHOSMTLYNSVSGELSMVLTGAFFEHKYAYLYSYVPOAIRMVDEYINCEDIAMNELY 840  
 DB 781 phosmtlynsvsgelsmvltgaaffehkyaylysyvpoairmvdeyincediamnely 840  
 QY 841 SHITRKPIKVTSRMTFRCPGCPQALSHDHSHERHKCNINFFVKYGYWPLIYTGFRVD 900

DB 841 shitrkpiikvtsrmtfrpcpgcpqalshdshsherkcninffvkgywpllytgfrvd 900  
 QY 901 SVLEKTRLPHPDKTCPKFI 919  
 DB 901 svlfekrtrlphtkctckfi 919

RESULT 2

AA84669  
 ID AAY84669 standard; Protein; 919 AA.

AC AAY84669;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of a human TRFX protein.

KW Tumour necrosis factor receptor-associated Factor; TRAF;  
 KW TRF-protein-interacting hereditary multiple extoses protein; TRFX;  
 KW signal modulator; tumour necrosis factor receptor;  
 KW CD40 mediated signal transduction; TRAF protein; cancer;  
 KW hereditary multiple extosis; autoimmune disease.

XX Homo sapiens.

PN W0200018959-A1.

PD 06-APR-2000.

PE 17-SEP-1999; 99WO-US21654.

PR 17-SEP-1998; 98US-0156191.

PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI Sato T;

DR WPI; 2000-293180/25.

DR N-PSDB; AAA12734.

PT New nucleic acid encoding Tumour necrosis factor Receptor-Associated  
 PT Factor (TRAF) protein-interacting hereditary multiple extoses (TRFX)  
 PT protein, useful in the diagnosing cancer -

PS Claim 39; Fig 8A-B; 161pp; English.

CC The present sequence represents a human tumour necrosis factor receptor-  
 CC associated factor (TRAF) protein-interacting hereditary multiple  
 CC extoses (TRFX) protein. TRFX is a signal modulator which bridges  
 CC between the tumour necrosis factor (TNF) receptor and CD40 mediated  
 CC signal transduction. Anti-TRFX antibodies are useful for treating an  
 CC abnormality in a patient by inhibiting binding of a TRFX protein and  
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality  
 CC is cancer, a hereditary multiple extosis or an autoimmune disease. The  
 CC cancer is colon cancer, gastric cancer, human head and neck squamous  
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,  
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,  
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,  
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.

XX Sequence 919 AA:

Query Match 100.0%; Score 4873; DB 21; Length 919;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGYTMLRNGAGNGGQCMLRMSNRIRLTWLSFTLVLFVLPPLIAHYLTTLDEADBA 60  
 DB 1 mtygmrlngagnggqctmlrwsnrirltwlsftlvlfvlppliahyltltldeada 60  
 QY 61 GKRIFGPRVGNELCEVKHVLDCRIRSVSEELLQLEAKROELNSEIAKLNLKIEACKS 120

```
Db 61 gkrlfsgprvneicevkhvldlcrilresveeellqleakrqelnsejaklnlkeacks 120
Oy 121 IENAKODLLQKKNVISOETESYKELMAQNOPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
Db 121 IENAKODLLQKKNVISOETESYKELMAQNOPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
Oy 181 NCFYSCPLTSGFPVYVSDQFVFGSYLDPLVKQAFQATARANYVTENADIACLYVI 240
Db 181 ncfyscpltsfgfprvyvsdqdfvfgsyldplvkqafqataranyvtenedaclyvi 240
Oy 241 LVGEMOPVVLRAPELEKOLYSLPHMRTDGNHVTILNSKSDPONLLYNVSTGRANVAQ 300
Db 241 lvgemopvvlrapelekylysipmwtcdgmhvtlinsksdpgnllynvstgranaq 300
Oy 301 STFYTVQYRRGFDLVSPLYHAMSEPNFMETIPQVPVKRYLFTFOGEKTESLRSIQEA 360
Db 301 stfytvgyrpgfdlvsplylhamsepnfmelpqvpvykrylftfgeketeslrsiqea 360
Oy 361 RSFEEMEGDPPADYDRIATATLKAQVDSKLDQVLEFTCKNOPKPSLPTEMALCGERD 420
Db 361 rsfeemegdpaddydrilatlatlkavqdsklqdvlfvftcknkpkslptemalcgere 420
Oy 421 RLELLKSTFALITTPGDPRLVSSGCATRLFEALEVGAAPVVLGEVOLPYODMLQNE 480
Db 421 rlellkistfallitpgdprlvssgcattrlfealevgavpvvlgevgqlpygdmqlwne 480
Oy 481 AALVPRPRVTEVHFLRLSLSDSLAMRQGRFLMETYESTADSIPTNYLAMIRRIQI 540
Db 481 aalvprprvtevnhflrlsldsdlamrqgrflmetystadsiptnylamirtriqi 540
Oy 541 PAAIRREMAAEPIHRSKGAAGTDPNMADNGDLDGVEYETPPYASRYLRNFTLYTDF 600
Db 541 paairreemaapeihrskaagtdpnmadngdldgveetppypasrylrnftlytdf 600
Oy 601 YRSNCAKAPGPHLPHPFPFVLPSEAKFLSGSGFRPIGAGSGSKECOALGHNVR 660
Db 601 yrsncakapghlphfpfpvlpseakflsgsgfrpiggagsgskecoalgahnvr 660
Oy 661 EQFTVMLTYEREELVNLNSLERLNGRLPYLNKVVVWNSPKLPSBDLMPDIGVIMVVT 720
Db 661 eqftvmltyereevnlmslerlngrlpylnkvvnwvnsplpsedllmpdigvimvvt 720
Oy 721 EKSNLNRPLPMNIEFETALISIDDAHLRHDIMEGFRVWRKADRIYVFGYRIAMDI 780
Db 721 eksnlnrplpmniefetalsiddahlrhdimegfrvwrkadrlyvfyryhawi 780
Oy 781 PHOSMLNSNSCELSMVLGAAFEHKKYAYLYSYVMPQALRDVDEYINCEDIAMNELV 840
Db 781 phosmlnsnscelsmvlgaafekhkyaylysyvmpqalrdvdeyincediameflv 840
Oy 841 SHITRRPPIKVTSMWTRPCGCPQALSHDSHFHERHKCNIFEVKKVGYMPLLYTQFRVD 900
Db 841 shitrppikvtsmwtrpcgcpqalshdsfhherhkcni fekvkygmpllytqfrvd 900
Oy 901 SVLFKTRLPHDKTRCEKFI 919
Db 901 svlfktrlpndktckfki 919

RESULT 3
AAB51341 standard: Protein: 919 AA.
XX
AC AAB51341:
XX
DT 05-APR-2001 (first entry)
XX
Rat Reg-binding protein SEQ ID NO:4.
XX
Rat: Reg-binding protein; Reg receptor; Reg: DNA synthesis; apoptosis;
KM cell proliferation; diabetes; antidiabetic.
XX
```

```
OS Rattus norvegicus.
PN WO200077192-A1.
XX 21-DEC-2000.
PD 09-JUN-2000; 2000MO-JP03764.
XX 10-JUN-1999; 99JP-0164488.
XX (OKAM/) OKAMOTO H.
PA Okamoto H;
PI Okamoto H;
XX WPI: 2001-061871/07.
DR N-PSDB; AAF26844.
XX
PT Protein binding to pancreatic Reg protein and its encoding DNA, useful
PT for screening candidate antidiabetic drugs
PS Claim 1; Page 74-81; 11pp; Japanese.
XX
CC The present invention describes DNA which encodes: (a) a protein of rat
CC origin binding to Reg protein; or (b) a Reg-binding protein derived from
CC the natural Reg sequence by addition, deletion and/or substitution of
CC one or more amino acid residues. The present sequence represents a rat
CC Reg-binding protein from the present invention. Reg-binding proteins
CC have antidiabetic activity. Drug compositions comprising Reg-binding
CC proteins and their genes are useful for the treatment and/or prevention
CC of diabetes. In addition, they are useful in developing new treatments
CC for diabetes.
XX
SQ Sequence 919 AA:

Query Match 97.6%; Score 4755; DB 22; Length 919;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 894; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

Oy 1 MTGYTMLRNGAGNGGOTCMLRMSNRITLWLSFTFLVILVFPPLAHVYLTLLDEADEA 60
Db 1 mtgytmlrngngnggctmlrwsnrilcwlstflilvlfpplahylyltcldeadea 60
Oy 61 GKRIFGPRVNEICEVKNHVDLCRIRESVSEBELQLEAKNOELNSETAKNLKIEACKS 120
Db 61 gkrlfsgpragneicevkhvldlcrilresveeellqleakrqelnsejaklnlkeacks 120
Oy 121 IENAKODLLQKKNVISOETESYKELMAQNOPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
Db 121 IENAKODLLQKKNVISOETESYKELMAQNOPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
Oy 181 NCFYSCPLTSGFPVYVSDQFVFGSYLDPLVKQAFQATARANYVTENADIACLYVI 240
Db 181 ncfyscpltsfgfprvyvsdqdfvfgsyldplvkqafqataranyvtenedaclyvi 240
Oy 241 LVGEMOPVVLRAPELEKOLYSLPHMRTDGNHVTILNSKSDPONLLYNVSTGRANVAQ 300
Db 241 lvgemopvvlrapelekylysipmwtcdgmhvtlinsksdpgnllynvstgranaq 300
Oy 301 STFYTVQYRRGFDLVSPLYHAMSEPNFMETIPQVPVKRYLFTFOGEKTESLRSIQEA 360
Db 301 stfytvgyrpgfdlvsplylhamsepnfmelpqvpvykrylftfgeketeslrsiqea 360
Oy 361 RSFEEMEGDPPADYDRIATATLKAQVDSKLDQVLEFTCKNOPKPSLPTEMALCGERD 420
Db 361 rsfeemegdpaddydrilatlatlkavqdsklqdvlfvftcknkpkslptemalcgere 420
Oy 421 RLELLKSTFALITTPGDPRLVSSGCATRLFEALEVGAAPVVLGEVOLPYODMLQNE 480
Db 421 rlellkistfallitpgdprlvssgcattrlfealevgavpvvlgevgqlpyndmlqne 480
Oy 481 AALVPRPRVTEVHFLRLSLSDSLAMRQGRFLMETYESTADSIPTNYLAMIRRIQI 540
Db 481 aalvprprvtevnhflrlsldsdlamrqgrflmetystadsiptnylamirtriqi 540
```

Db 481 aalvvpkprvtevhflrlsdsdillamrgrflwetyfstadsifntlamitrtiqi 540  
 QY 541 PAAPIREEAAAEIPHRSGKAAGTDPNADNGDLGCVETEPYPYASPRYLRNFLLVTDF 600  
 Db 541 PPAPIREEAAAEIPHRSGKAAGTDPNADNGDLGCVETEPYPYASPRYLRNFLLVTDF 600  
 QY 601 YRSMNCAFGPFLPFPHTPPDPLPSEAKFISGTFGRPRIGGAGSGKEFOALGNPR 660  
 Db 601 YRSMNCAFGPFLPFPHTPPDPLPSEAKFISGTFGRPRIGGAGSGKEFOALGNPR 660  
 QY 661 EGFVVMILYEREVEVLMSLERLNGLPYLKVVVWVNSPKLPSEDLMPDVGPIWVPT 720  
 Db 661 EGFVVMILYEREVEVLMSLERLNGLPYLKVVVWVNSPKLPSEDLMPDVGPIWVPT 720  
 QY 721 EKNSLNRRFLPKNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHADI 780  
 Db 721 EKNSLNRRFLPKNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHADI 780  
 QY 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840  
 Db 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840  
 QY 841 SHITRRPPIKYSRMTFRCPCGPQALSHDSDHFERHAKCINFEVYVYMPPLTYQFVND 900  
 Db 841 SHITRRPPIKYSRMTFRCPCGPQALSHDSDHFERHAKCINFEVYVYMPPLTYQFVND 900  
 QY 901 SVLEKTRLPKHTKCFKFI 919  
 Db 901 SVLEKTRLPKHTKCFKFI 919

## RESULT 4

AAY84668  
 ID AAY84668 standard; Protein; 918 AA.

AAV84668;

25-JUL-2000 (first entry)

Amino acid sequence of murine TREX protein.

Tumour necrosis factor receptor-associated Factor; TRAF;

TRF-protein-interacting hereditary multiple extoses protein; TREX;

signal modulator; tumour necrosis factor receptor;

CD40 mediated signal transduction; TRAF protein; cancer;

hereditary multiple extosis; autoimmune disease.

Mus sp.

WO200018959-A1.

06-APR-2000.

17-SEP-1999; 99WO-US21654.

17-SEP-1998; 98US-0156191.

(UYCO ) UNIV COLUMBIA NEW YORK.

Sato T;

WPI; 2000-293180/25.

N-PDB; AAA12733.

New nucleic acid encoding Tumour necrosis factor Receptor-Associated

Factor (TRAF) protein-interacting hereditary multiple extoses (TREX)

protein, useful in the diagnosing cancer -

Claim 38; Fig 7A-B; 161pp; English.

CC between the tumour necrosis factor (TNF) receptor and CD40 mediated  
 CC signal transduction. Anti-TREX antibodies are useful for treating an  
 CC abnormality in a patient by inhibiting binding of a TREX protein and  
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality  
 CC is cancer, a hereditary multiple extosis or an autoimmune disease. The  
 CC cancer is colon cancer, gastric cancer, human head and neck squamous  
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,  
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,  
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,  
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.  
 CC  
 XX Sequence 918 AA:

Query Match 96.9%; Score 4720.5; DB 21; Length 918;  
 Best Local Similarity 97.0%; Pred. No. 0;  
 Matches 891; Conservative 7; Mismatches 20; Indels 1; Gaps 1;  
 QY 1 MTGYMLRNGANGNGOTCMRLMSNRRLTWLSFTLVILVFPPLIAHYVLTIDEADEA 60  
 Db 1 mtygmrlngngvgnpgqcmrlmsnrirltwlsftllvffpliahyytltddeadea 60  
 QY 61 GKRIFGPRVGNLCEVKNHVLDCRIRSVSEELLQLEAKRQELNSETAKLMTKEACKS 120  
 Db 61 gkrifgprvgnlcevknhvldcrrsvseellqleakrqlnsetaklnlkeacks 120  
 QY 121 IENAKODLLQKNNYSQTEHSYKELMAONPKLSIPRLPEKNDAGLPPEKATRGCLH 180  
 Db 121 ienakodllqknnysqtehsykelmagngklsiprllpekndaglppekavrrgclh 180  
 QY 181 NCFDYSRCPLTSGFPVYVYDSQFVFGSYLDPVKQAFOATARANVYVENADIACLYVI 240  
 Db 181 ncfdysrcpltsghfpvyvysdqfvgfsgylpdkvqafataranvyvenadiaclyvi 240  
 QY 241 LVGMEQEPVYLRPALEKOLYSLPHMRPDGHNHITINLSKSDONQLNYSSTRANVAQ 300  
 Db 241 lvgmeqepvylrpalekolslphmrpdgghnhitinsksdonqlnysstranvaq 300  
 QY 301 STFYVOYRPGFDLVSPVHAMSEPNMEIPPOVPKRYLFFQGEKTESLRSSLOEA 360  
 Db 301 stfyvoyrpgfdlvspvhamsepnmeipovpkrylffqgeketeslrssloea 360  
 QY 361 RSFEENBGPDPADYDRIATLKAVDSDKIDQVLEFTECKNQKPSLPTEWALGGERED 420  
 Db 361 rsfeenbgpdpadydriatlkaavsdskidqvleftecknqkpslpewalgered 420  
 QY 421 RLELKLSTFALITTPGDPRLVYSSGCATRLFEALFEGAVPVYGEVOVLPYQDMQWNE 480  
 Db 421 rlelklstfaliitpgdprlvysgcatrlfealfegavpvylgevoqlpydmqlwne 480  
 QY 481 AALVVPKPRVTEVHFLRSLSDSDDLAMRRGRFLMETVSTADSIPTVLIAMIRFIQI 540  
 Db 481 aalvvpkprvtevhflrlsdsdillamrgrflwetyfstadsifntlamitrtiqi 540  
 QY 541 PAAPIREEAAAEIPHRSGKAAGTDPNADNGDLGCVETEPYPYASPRYLRNFLLVTDF 600  
 Db 541 PPAPIREEAAAEIPHRSGKAAGTDPNADNGDLGCVETEPYPYASPRYLRNFLLVTDF 600  
 QY 601 YRSMNCAFGPFLPFPHTPPDPLPSEAKFISGTFGRPRIGGAGSGKEFOALGNPR 660  
 Db 601 YRSMNCAFGPFLPFPHTPPDPLPSEAKFISGTFGRPRIGGAGSGKEFOALGNPR 660  
 QY 661 EGFVVMILYEREVEVLMSLERLNGLPYLKVVVWVNSPKLPSEDLMPDVGPIWVPT 720  
 Db 661 EGFVVMILYEREVEVLMSLERLNGLPYLKVVVWVNSPKLPSEDLMPDVGPIWVPT 720  
 QY 721 EKNSLNRRFLPKNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHADI 780  
 Db 721 EKNSLNRRFLPKNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHADI 780  
 QY 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840  
 Db 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840

Oy		841	SHITKPKPIKTYTSMTPTFCPCPGCALSHDDSHFERHRHNCINFVAVYGYMPLLTQGRVD	900
Dd		840	shltkpkpiKysrvtfcpgcpgalshddshfherhkcnlfvkvgympillycgfrvd	899
Oy		901	SVLFTKRLPHDKTKCFKI	919
Dd		900	svlftkrlphdktkcfki	918
RESULT	5			
ID	ABB70158			
AB	ABB70158	standard; Protein:	972 AA.	
AC	ABB70158;			
AD				
AE				
AF				
AG				
AH				
AI				
AJ				
AK				
AL				
AM				
AN				
AO				
AP				
AQ				
AR				
AS				
AT				
AW				
AX				
AY				
AZ				
BA				
BB				
BC				
BD				
BE				
BF				
BG				
BH				
BI				
BJ				
BK				
BL				
BM				
BN				
BO				
BP				
BQ				
BR				
BS				
BT				
BU				
BV				
BW				
BX				
BY				
BZ				
CA				
CB				
CC				
CD				
CE				
CF				
CG				
CH				
CI				
CJ				
CK				
CL				
CM				
CN				
CO				
CP				
CQ				
CR				
CS				
CT				
CU				
CV				
CW				
CX				
CY				
CZ				
DA				
DB				
DC				
DD				
DE				
DF				
DG				
DH				
DI				
DJ				
DK				
DL				
DM				
DN				
DO				
DP				
DQ				
DR				
DS				
DT				
DU				
DV				
DW				
DX				
DY				
DZ				
EA				
EB				
EC				
ED				
EE				
EF				
EG				
EH				
EI				
EJ				
EK				
EL				
EM				
EN				
EO				
EP				
EQ				
ER				
ES				
ET				
EU				
EV				
EW				
EX				
EY				
EZ				
FA				
FB				
FC				
FD				
FE				
FF				
FG				
FH				
FI				
FJ				
FK				
FL				
FM				

QY	105	SEIAKLNLKTEACKSCKSIENAKODLLQKLVNISTQHSYKVELMAQNOPKLSLPIRLPEKD	164
Db	137	sdlsqgkleeelkqellreqtelrklisvegaqvaqeaavgrntpdlaiprsllpntc	195
QY	165	DAGLPPPK-----ATMGCRLLHNFQDSRQPLTSGPPVAVYDSDQVVF-----GSYLDPLVK	215
Db	196	---lprksnpltgmaascemhncfnmsrctslsgfipvlyldpddensvgrkgydidgflk	252
QY	216	QAFQATARANYVYVEMNDADICLVYIVTEG--MOEPVVLK-----	252
Db	253	tlkqtlgynahaiykkdkhacelylvtlgeaelldqdlrmryaaqaeehqspetclend	312
QY	253	-PAELEKQOLSLEPMRTDGNHNVIIINSRK---SDQNLILNVSTGRAMVAOSTFYVQY	308
Db	313	cpvmdnek-lyglilpwyggdgdnrhvlllnlarcdltshtnplryqntmraltvgsafereqf	371
QY	309	RGPFLVAVSPILVHAMSDEPNMEETIPQVYVARKTLTFPGQEKLESLSSIQEARSFEEME	368
Db	372	tpgydllyvpplilpppgdgdwqecaemparckyllytqge-lrpksaslnpdaillenh	430
QY	369	GDPPEADYDRIATILKAVNODSLDQVLYVEFTC-----KNQKPSLPEMALGKEREDRLLEL	424
Db	431	ad-----makgatdqdfqlqdcvqateqegedslp-dwlcgsdsasrql	475
QY	435	LKLSTFALITPGDPLVYISGCATRLFEALEYGAVPVVLG--EQVOLPYQMLQWNEAL	483
Db	476	lkdsffslilppllngr-vsstlmlarlyeealrsgavpavlgedelilpyleavdwrtel	534
QY	484	VYPKRYVEVFFLRSISDSQSLAMRQGFLEMETFSTADSLFNTVLAMITRRQIOPA	543
Db	535	llpkartlelflilraygdadalllllrrqglllweylyssvqavdcvtlslltdrlgiprr	594
QY	544	PIREEAAE-----IPHRSGKAAGTDPPNNAADGDLGLGPETEPPEYASPRYLNRFTLV	597
Db	595	pvpavlaqsvfnstfflrlksdprvgldtepees---lgrl--eppypapafirnyfllr	648
QY	558	TDFYISWACGAPRHLPHPHPDVPVLSEKFLGSGTGPRTTGGAGSGSGKEFOALGCGN	657
Db	649	mqaekaeawndlpfilylpqrlpfdpalsreakfmgshntgfrp19kyl99agkefgeal99n	708
QY	658	VPRQFTVVMVLTREEEKLMSLERLNGLPYLKKNVVMVNSPKLPSEDLMPDIDGVPWV	717
Db	709	ypreqfclvmltyereqylngslgrlygdyllkvvvvvnspkprpiddlrrpddlyprav	768
QY	718	VRTKNSLNNRFLPMNETETETALISIDDOAHLRHDEIMEGFRVWRBARDRIVGPGRYHA	777
Db	769	lrapsnslnnrlpfldveteaavsvdadahlrhndellfgfwreghrdgrvvgfprgyna	828
QY	778	WDI--PHOSMLYNSNYSCELSMVLTGAAFRHKYATLYLSVMQALRDWVDEINCEDIA	835
Db	839	wdlngpqwqynhsnyscealsmvtlgeaaivhkykyllylylvhlpqalrtldvdeimcedla	888
QY	836	MNFIATSHTRKRPPIKIVYSRMTFRCPGQALSHDSDHFHEHNKICINFPAKVVGYMPLLYT	895
Db	889	mnflvshlrrppkvksasrwtfcfpgspvslseddchfgerhcklnffsrfvgypllnt	948
QY	896	QFRDVSULEKTRLPDXTKCFKFT	919
Db	949	qyradsllfktrrlphdkqckfky1	972
RESULT 6			
AAB51340			
ID AAB51340 standard; Protein: 364 AA.			
XX AAB51340;			
XX DT 05-APR-2001 (first entry)			
XX Rat Reg-binding protein SEQ ID NO:2.			
DE Rat Reg-binding protein; Reg receptor; Reg; DNA synthesis; apoptosis;			
XX Rat; Reg-binding protein; Reg receptor; Reg; DNA synthesis; apoptosis;			



cell proliferation; diabetes; antidiabetic.  
Rattus norvegicus.  
W0200077192-A1.  
21-DEC-2000.  
09-JUN-2000; 2000WO-JP03764.  
10-JUN-1999; 99JP-0164488.  
(OKAMOTO) OKAMOTO H.  
Okamoto H;  
WPI; 2001-061871/07.  
N-PSDB; AAF26843.  
Protein binding to pancreatic Reg protein and its encoding DNA, useful for screening candidate antidiabetic drugs -  
Claim 1; Page 61-64; 11pp; Japanese.  
The present invention describes DNA which encodes: (a) a protein of rat origin binding to Reg protein; or (b) a Reg-binding protein derived from the natural Reg sequence by addition, deletion and/or substitution of one or more amino acid residues. The present sequence represents a rat Reg-binding protein from the present invention. Reg-binding proteins have antidiabetic activity. Drug compositions comprising Reg-binding proteins and their genes are useful for the treatment and/or prevention of diabetes. In addition, they are useful in developing new treatments for diabetes.  
Sequence 364 AA;  
Query Match 34.2%; Score 1666.5; DB 22; Length 364;  
Best Local Similarity 85.4%; Pred. No. 9.8e-150;  
Matches 323; Conservative 9; Mismatches 19; Indels 27; Gaps 2;  
1 MTGTMRLNGAGNGGOTCMRLMSNRIRLMTLSPTLVIIIVFPFLIAHYITLDEADEA 60  
1 mtygtmlrngvgnggqctmlrwmnrirlltwistlflilvfpfliahyytlcdeaada 60  
61 GKRIIFGRVGNELCEVHVLDLCRIRSVSEELQLEAKROELNSETAKNLKTEACKS 120  
61 gkriifgrvgnelcevhvldlcrirsvseellqlleakrqelnselaklnlkeacks 120  
121 IENAKODLQKVNISTEHSYKELMAONQPKLSLPIRLLEPKDDAGLPPKATRGCRH 180  
121 ienakodlqlkvnistehsykelmaonqpklsiprllpekdaglppkavtrgrl 180  
181 NCFDYSRCPLTSGFPVYVYSDQFVSGYLDPLVYKQAFQATARANYVTENADACIYI 240  
181 ncfdysrcpltsqfpyvyysdqfvgysldplvlykqafqatvanyvtenaadaclyyv 240  
241 LVSEMOEPVVLRAPELEKOLYSLPHWRTDGHNVYINLSKSPDONLILYVSTGRVAAQ 300  
241 lvsemoepvvlrapelkolyslphwrtldghnvynlslskspdonlilyvstgrvaaq 300  
301 STEYTVQYRGGFDLVSPVHAMSEPNFMEIPQVPYKRYLFTFOGEKTESLRSSLOEA 360  
301 steytvqyrggfdlvspvhamsepnfmeipqovpykrylftfogekteslrssloea 360  
361 RSFEEME-----GDPP 372  
361 rsfeeme-----gdpp 372  
340 grgeeeqeklgvwrgrpp 357  
340 grgeeeqeklgvwrgrpp 357  
RESULT 7  
AA84675  
ID AAY84675 standard; Protein; 262 AA.

AA84675;  
25-JUL-2000 (first entry)  
Amino acid sequence of a protein.  
Tumour necrosis factor receptor-associated Factor; TRAF;  
TRF-protein-interacting hereditary multiple exostos protein; TREX;  
CD40 mediated signal transduction; TRAF protein; cancer;  
hereditary multiple exostos; autoimmune disease.  
Unidentified.  
W0200018959-A1.  
06-APR-2000.  
17-SEP-1999; 99WO-US21654.  
17-SEP-1998; 98US-0156191.  
(UYCO ) UNIV COLUMBIA NEW YORK.  
Sato T;  
WPI; 2000-293180/25.  
New nucleic acid encoding Tumour necrosis factor Receptor-Associated Factor (TRAF) protein-interacting hereditary multiple exostos (TREX) protein, useful in the diagnosis cancer -  
Disclosure; Page 150; 161pp; English.  
The specification describes human and murine tumour necrosis factor receptor-associated factor (TRAF) protein-interacting hereditary multiple exostos (TREX) protein. TREX is a signal modulator which bridges between the tumour necrosis factor (TNF) receptor and CD40 mediated signal transduction. Anti-TREX antibodies are useful for treating an abnormality in a patient by inhibiting binding of a TRAF protein and a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality is cancer, a hereditary multiple exostos or an autoimmune disease. The cancer is colon cancer, gastric cancer, human head and neck squamous cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer, oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer, papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma, giant cell tumour, Ewing sarcoma, or other malignant tumours. The present sequence appears in the specification.  
Sequence 262 AA;  
Query Match 29.5%; Score 1436; DB 21; Length 262;  
Best Local Similarity 100.0%; Pred. No. 5.4e-128;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
658 VPPROFVWVLTREBERVNLNLSLEPLNGLPYLKVVVWVMSPLPSBDILMPDIGYPIW 717  
658 vpprofvwvltrebervnl nlsleplnglpylkvvvwmvmsplpsbdilmpdigypiw 717  
1 vpreqitvmllyereevlmslerlmglylnkvvvwmvmspxlpsedilmpdiqyplmw 60  
1 vpreqitvmllyereevlmslerlmglylnkvvvwmvmspxlpsedilmpdiqyplmw 60  
718 VREKSNLNNRFLPMNEIETEAISIDDDAHLHDEIMFGFRVWRARDRIVFPGRYHA 777  
718 vreksnlnnrflpmnei eteaisidddahlhdeimfgfrvwrardrivfpgryha 777  
61 vrtkxslmrflpweieeteaisidddahlhdeimfgfrvwrardrivfpgryha 120  
61 vrtkxslmrflpweieeteaisidddahlhdeimfgfrvwrardrivfpgryha 120  
778 WDIPHOSWLYNSVYSCSLSVLTGAAPFHKYAYLYSYVWPAQIRDMVDEYINCEDIAMN 837  
778 wdiphoswlynsyvsclsvltgaapfhkyaylysyvwpaqirdmvdeyincediamn 837  
121 wdlphqswlynsyvsclsvltgaafhkyaylysywmpqairdmvdeyincediamn 180  
121 wdlphqswlynsyvsclsvltgaafhkyaylysywmpqairdmvdeyincediamn 180  
838 FLVSHITRKPRIVVTSRWMTFRGCGPQALSHDSHHERHCKINFPVKKYGYVPLIXTQF 897  
838 flvshitrkprivvtsrwmtfrgcgpqalshdshherhckinfpvkkyygyvplixtqf 897  
181 flvshitrkprivvtsrwmtfrgcpqalshdshherhckinfpvkkyygywpllytqf 240  
181 flvshitrkprivvtsrwmtfrgcpqalshdshherhckinfpvkkyygywpllytqf 240

QY 898 RVDVLFKTRLPDHTKCEKFI 919  
DB 241 RVDVLFKTRLPDHTKCEKFI 262

RESULT 8

ID AAY28883 standard; Protein: 718 AA.

AC AAY28883;

DT 01-FEB-2000 (first entry)

DE Human Ext-2 protein.

KM Human Ext-2 protein; Ttv protein; hedgehog interacting protein;  
KW protein core; proteoglycan; hedgehog-mediated signal transduction;  
KW hedgehog specific GAG chain; Ext dependant pathway; drug discovery assay;  
KW growth factor; apoptosis; cell proliferation; cell differentiation;  
KW cancer.

OS Homo sapiens.

PN MO950385-A2.

PD 07-OCT-1999.

PF 30-MAR-1999; 99MO-US06892.

PR 30-MAR-1998; 98US-0079928.

PA (HARD ) HARVARD COLLEGE.

PI Bellaiche Y, The SI, Perlimon N;

DR WPI: 1999-610841/52.

DR N-PSDB; AA208224.

PT Identifying inhibitors of signal transduction of extracellular  
PT proteins, used for developing agents for treating, e.g. proliferative  
PT disorders or cancers or for cartilage -

PS Claim 9; Page 103-105; 141pp; English.

CC The present sequence is human Ext-2 protein. It is 25% identical to Ttv  
CC protein. Exts are hedgehog interacting proteins which bind to hedgehog  
CC protein with high affinity to regulate diffusion. They are a family of  
CC glycosyltransferases synthesizing GAG chains attached to the protein core  
CC of proteoglycans. Hedgehog-mediated signal transduction can be inhibited  
CC by inhibiting the synthesis of hedgehog specific GAG chains by Ext  
CC dependant pathways. Drug discovery assays identifies such agents that  
CC modify the biological activity of growth factors by inhibiting Ext  
CC activity to prevent hedgehog-dependent proliferation of cells. This is  
CC useful in treating disorders related to aberrant apoptosis, cell  
CC proliferation or differentiation, cancer etc.

SO Sequence 718 AA;

Query Match 20.5%; Score 1000; DB 20; Length 718;  
Best Local Similarity 32.5%; Pred. No. 1,1e-85;

Matches 260; Conservative 155; Mismatches 258; Indels 156; Gaps 27;

QY 140 HSKELMAQNPRLST--PIRLPERKDDGLP-PPKATGCRLLHNCFDYSRCPPLTSGF-194  
DB 48 hskelmaqnpvrlstpyvrlp-----adspipergdlsctmhtctfdyrc-----gfn 99

QY 195 -----PYVYDSQGF--FGSYLDPLVKQAFQATARA---NVYVENADIACLVILLGE 244  
DB 100 pknkikyalyalkkyvddfgvsvantlsreyneilmalsddytdinacilfypsldv 159

QY 245 MQEYVVRAPAELEKQLSLPHWRTDGHNIYIN-LSKRSOTQNLTVNSYGRANVAQSTF 303

DB 160 lngn-flrlketagmagjstrw-drglnhllfmlpggppdyntaldvprdrailagggf 217  
QY 304 YTVQYRPGFDLVVSPVYHAMSEPNFMEIPQVAVKRRYLTTFQGEKIESIRSSLSQBARSP 363  
DB 218 stwttyrgydvsl-pyysplsaee-vdlpekypgpygyl-----lsgyqghpey 265

QY 364 EEEEMGDPPADYDRIATLKAQVDSKLDQVLEFTCKNPKPSLPTMALGGERE--DR 421  
DB 266 redle-----alqvkhgeavlvldctnlsegvlsvr-krchkhqyfdy 308

QY 422 LELLKSTFALITTPDPRVVISGCGATRLFEALENGAVPVVLEGOVQLPYOMLQNEA 461  
DB 309 pylvgeatfcvvlr--garl---ggav-lsdvlgagcvpvladsyilpfsevlwkra 361

QY 482 ALVVPKPRVVEHFLRLSDSLAMRQGRFLMETYFSTADSIPTVYAMIRTRIQP 541  
DB 362 svvvppekmdysllqslprqleemqrqarvfwesfslakalatlqindrtlypy 421

QY 542 AADIREBAAEIRHRSKGAAGTDPNMADGDLGPEVTEPPYASPRYLNFLLVTDYF 601  
DB 422 aalsyee-----wmdpav----- 435

QY 602 RSMNCAPGRPHLPDPDPLPSEAKFLGSGTGFRRIGGAGSGSGERQALGYNPRE 661  
DB 436 -kwgsvsnpilf-----plppqs-----q 454

QY 662 QFTVVMILTYREBEVLMNSLERLNGDLPYLNKVVVWVWNSP-KLPSEDLMPDIGVIMVRT 720  
DB 455 gftaivltgydrveslffvilevskypslskllvwmnqknpedalswplrvlpykvtc 514

QY 721 EKNSLNRLPWNIEETEAISTDD-ANLRHDEIMGFVWREARDRIVGPRGYHAWD 779  
DB 515 aenklsnrffpydeleavlaiddlmltsdelqfygvevrefprlvygryhlw 574

QY 780 IPHQSWLYNSNVCESLNVLTGAAPFKKYAYLYSYVMPAIDMDVEYINCDIANNFL 839  
DB 575 hemkwysewlewsmtvltgaafykhkyfnylytkmpgdlkhwvdaImnncedlmnfl 634

QY 840 VSHTRKPRPIKVTSMRTFRQCP--QALSHDSHFHERKCNFVKVGYVPLLYTQF 897  
DB 635 vanvtgkavlkvprkfkfpectaldglsldqtmwresclnkfsavfgtmplkveh 694

QY 898 RVDVLFKTRLPDHTKCF 916  
DB 695 radpvlykddfp-eklsf 712

RESULT 9

ID AAB58946 standard; Protein: 663 AA.

AC AAB58946;

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 654.

KM Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KM neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KM antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;  
KM antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
KM Addison's disease; allergy; autoimmune haemolytic anaemia;  
KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KM cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.

PN MO200055173-A1.

PD 21-SEP-2000.





```

Db 320 -----ekydyremhlnatfcly-prgrrl-----gsfrfleaigaacypvmlnsngwel 366
Qy 471 PYODMLQNMALVYPRKPRTEVHFLRLSDSDLLAMRQGFMEYETSTADSTFENVY 530
Db 367 pfsewvnmgaavlgderlilqstcrlshqkllalqgtflweayfssvekyltlt 426
Qy 531 LAMITRTIOIPAAPIREBAAEPIHRSRGAAGTDPNMDNGDLDPVETEPYASPRYL 590
Db 427 leiigdrf-----fkhis----- 439
Qy 591 RNFETLVDFYRSMNCAPCPHILFPHTPDPVLPSEAKFLG-----SGTGRPTGGAG 644
Db 440 rnsli-----wnkhpq--glf-----vlpqysylgdfpyyanlglkpr----- 476
Qy 645 GSGKEPQALGQNPREQTVMATYEREVEYLMNSLERLNGLPYLANKVYVWVNSPK-LPS 703
Db 477 --pskftaviahvcp-----lvsqspvilkilvaaksqcagllvncdpkrlpa 525
Qy 704 EDLMPDVGPIVMVYTERKNSLNRFLEPWEIETEALISIDDAHLRHDEIMGFVRWE 763
Db 526 kh-rwpatavpvyvlegesksrflpydnaltdavlsidedvtalstevfafatwqs 584
Qy 764 ARDRVGPGRYHNDIPHQSWLYNSNYSCELSMVLGAAFFHKYAYLYSYVMPQAIRD 823
Db 585 fperivgyparshidnakerwgltskwtndysmvltgaaalyhkyhylyshylypaslkn 644
Qy 824 MWDEXINCEDIAMNPLVSHITRKPIKYT-----SRMTFRCPGCPQALS 867
Db 645 mvdqqlancedlmlnflvaavtklppikvqlkqykyltmngqtsraswa----- 693
Qy 868 HDDSHFHRKCNINEFVKYGYMPLLYTQFRVDSVLFKTRL 908
Db 694 -dpdhiagqscmnlfaswfgymplhsqmdlpvlfkdyv 733

RESULT 12
AAy84680
ID AAY84680 standard; Protein: 259 AA.
XX
AC AAY84680;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of a protein.
XX
KM Tumour necrosis factor receptor-associated Factor; TRAF.
KM TRF-protein-interacting hereditary multiple extoses protein; TREX.
KM signal modulator; tumour necrosis factor receptor;
KM CD40 mediated signal transduction; TRAF protein; cancer;
KM hereditary multiple extosis; autoimmune disease.
XX
XX Unidentified.
XX
PN WO200018959-A1.
XX
PD 06-APR-2000.
XX
PF 17-SEP-1999; 99WO-US21654.
XX
PR 17-SEP-1998; 98US-0156191.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Sato T;
XX
DR WPI; 2000-293180/25.
XX
PT New nucleic acid encoding Tumour necrosis factor Receptor-Associated
PT Factor (TRAF) protein-interacting hereditary multiple extoses (TREX)
PT protein, useful in the diagnosing cancer -
XX
PS Disclosure; Page 155; 161pp; English.
XX

```

```

CC The specification describes human and murine tumour necrosis factor
CC receptor-associated factor (TRAF) protein-interacting hereditary multiple
CC extoses (TREX) protein. TREX is a signal modulator which bridges
CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
CC signal transduction. Anti-TREX antibodies are useful for treating an
CC abnormality in a patient by inhibiting binding of a TREX protein and
CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
CC is cancer, a hereditary multiple extosis or an autoimmune disease. The
CC cancer is colon cancer, gastric cancer, human head and neck squamous
CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
CC giant cell tumour, Ewing sarcoma, or other malignant tumours. The
CC present sequence appears in the specification.
CC
SQ Sequence 259 AA:

Query Match 17.4%; Score 848.5; DB 21; Length 259;
Best Local Similarity 58.1%; Pred. No. 5.4e-72;
Matches 151; Conservative 44; Mismatches 62; Indels 3; Gaps 2;

Qy 660 RQGTVMVLTYEREVEYLMNSLERLNGLPYLANKVYVWVNSPKLPSSDLMPLDGVPIVMYR 719
Db 3 rqtftvvltyerdaavgalerlmglypnlklyvwmvnrddpd-swpslhpveflr 61
Qy 720 TERNSLNRFLEPWEIETEALISIDDAHLRHDEIMGFVRWEARDRVGPGRYHAMD 779
Db 62 vaenlnrfypwdrleaeavlsiddidlmqgellafvrerndrtivgprahha-- 119
Qy 780 IPHQSWLYNSNYSCELSMVLGAAFFHKYAYLYSYVMPQAIRDMWDEXINCEDIAMNPL 839
Db 120 rygdsamlynshlcoqsmnlcgaaalhknlyltaylyempaelrehvnsikncediannyl 179
Qy 840 VSHITRKPIKYTSRMTFRCPGCPQALSDDSHEHHRKCNINEFVKYGYMPLLYTORRV 899
Db 180 vshltrkppiktsrwtlkcptcteslykgthfektnecmrlfcklygnplktsqtra 239
Qy 900 DSVLEFKTRLPHDKTKCFKEI 919
Db 240 dsllfkrtrlpqhkgkcfkyv 259

RESULT 13
ABb61749
ID ABb61749 standard; Protein: 717 AA.
XX
AC ABb61749;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide seq ID NO 12039.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05852.

```



```

Db      89  cmetctdfttc--ydrflvylypp-----eplnslgaaptsanyqklltaiges 137
Qy      227 -YVTENADIACLYVILVGENQEPVLRP--AELEKOLYSLPHMRTGHNHVTINLSRKS- 282
Db      138 ryltsdptaacilfvlgidtdrdsisedyrrnvpstlarlpw-nngnrhllfhllysgtlw 196
Qy      283 -DTQNLILYNSVSGRAMVAOSTFYTVQVRPGFDLVSPLVH-----AMSEPNFME 330
Db      197 pdyaenslfgdaegaallakasmgvlqrlngfdvsl-plfhkqglpragatgvtqsnmf-- 253
Qy      331 IPPQVPYKRRKYLFTFOGEKI-----ESLSSSQEARSFEEMEGDEPPADYDRIATLTL-- 383
Db      254 -----pankyllylafkkyrvhngisetrnslfhlhngf-----dmvlyltcor 296
Qy      384 -----KAVODSKLDQVLEVEFTCKNOKPRLPTFMALGGEREDLELLKSTFLALITTPGD 438
Db      297 hqkswrelqdnrcdednrey-----dryyelllqnstfcly--prg 336
Qy      439 PRLVISSGCATRLFEALEVGAVPVVLGEOVOLPYQDMLQWNEALVVPKRYTEVHFLRL 498
Db      337 rrl-----gsfrlealagagcipvlslsnawvpfeskikwkaalwaderilllqypdlvr 391
Qy      499 SLSDSDLLAMRQGRFLWETFEYSTADSIENFVTLAMIRTRIQIPAPPIREAAAEIPIHNSG 558
Db      392 sipaerifalrqtqylweryfgsiekivftfeller--lpdyprvssly----- 441
Qy      559 KAAGTDPNMADNGDDLGPVETEPYASPRYLRLNFTLVTDYFRSMNCAPGPFHLFP-HT 617
Db      442 -----wsspgalllprfta 456
Qy      618 PFDVYLPSEAKFLVSGTGFRIIGGAGSGKEFQALAGNVPREGFTVMLTYEREVYLM 677
Db      457 dsarymp-----flinsmgaeprhnytaivlyqgalgna-----aly 496
Qy      678 NSLERLNLGFLPYNTVYVYVWNSPK-LPSEDLMPDIOGPIVIMV-----RT 720
Db      497 klvtcticksgfverililwaadrpdlpkwtpscshlpilhlisgstrsgaagtsqct 556
Qy      721 E-KNSLNNRFLPWNIEITEALISIDDAHLRHDEIMGFGRVWRREARDIVGFPRYHAMD 779
Db      557 egrpslsgqrlfydeigtdavlsidedaillntdeldfaytwrtdfperivgypparahfwd 616
Qy      780 IPHOSMLYNSVSCLESMVLTGAAFHFHKKYAYLYSYVPOAIRDMVDEYINCEDIANNFL 839
Db      617 dsknawytickwlnyysilvtgaafhnylyltnwlslllktvgssncedilmlml 676
Qy      840 VSHITRKPPIKVTSRMFFRCPCQALSHDSDHHERHKCINFEVKYVGYMPLLYTQFRV 899
Db      677 vshvtrckprikvqtrglqsgdgsfawnopdhifqscntfaavfgympilrnlm 736
Qy      900 DSVLEFK 905
Db      737 dpmlyr 742

```

```

PD      27 -SEP-2001.
PF      23 -MAR-2001; 2001WO-US09231.
XX      23 -MAR-2000; 2000US-191637P.
PR      11-JUL-2000; 2000US-0614150.
XX      (PEKE ) PE CORP NT.
PI      Venter JC, Adams M, LI PMD, Myers EW;
DR      WPI; 2001-656860/75.
XX      N-PSDB; ABL08984.
PT      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
PS      Disclosure; SEQ ID NO 21435; 21pp + Sequence Listing; English.
CC      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      Sequence 760 AA;
SQ

```

Query Match 16.3%; Score 795.5; DB 22; Length 760;  
 Best Local Similarity 26.8%; Pred. NO. 3.7e-66;  
 Matches 227; Conservative 132; Mismatches 282; Indels 205; Gaps 26;

```

Qy      122 ENAK-----QDLQKWNYSOTENSHYKELMAQONPK-LSLPIRLPEKDDAGLPPKATRG 176
Db      40 esakdgvgvqhneqlpsfrfigahdmgelqllgngskjsdskslhlytrpdl----- 88
Qy      177 CRLNCFDYSRCPLTSGFPPVYVSDQFVFGSYIDPLVKAQFATARANV----- 226
Db      89 cmetctdfttc--ydrflvylypp-----eplnslgaaptsanyqklltaiges 137
Qy      227 -YVTENADIACLYVILVGENQEPVLRP--AELEKOLYSLPHMRTGHNHVTINLSRKS- 282
Db      138 ryltsdptaacilfvlgidtdrdsisedyrrnvpstlarlpw-nngnrhllfhllysgtlw 196
Qy      283 -DTQNLILYNSVSGRAMVAOSTFYTVQVRPGFDLVSPLVH-----AMSEPNFME 330
Db      197 pdyaenslfgdaegaallakasmgvlqrlngfdvsl-plfhkqglpragatgvtqsnmf-- 253
Qy      331 IPPQVPYKRRKYLFTFOGEKI-----ESLSSSQEARSFEEMEGDEPPADYDRIATLTL-- 383
Db      254 -----pankyllylafkkyrvhngisetrnslfhlhngf-----dmvlyltcor 296
Qy      384 -----KAVODSKLDQVLEVEFTCKNOKPRLPTFMALGGEREDLELLKSTFLALITTPGD 438
Db      297 hqkswrelqdnrcdednrey-----dryyelllqnstfcly--prg 336
Qy      439 PRLVISSGCATRLFEALEVGAVPVVLGEOVOLPYQDMLQWNEALVVPKRYTEVHFLRL 498
Db      337 rrl-----gsfrlealagagcipvlslsnawvpfeskikwkaalwaderilllqypdlvr 391
Qy      499 SLSDSDLLAMRQGRFLWETFEYSTADSIENFVTLAMIRTRIQIPAPPIREAAAEIPIHNSG 558
Db      392 sipaerifalrqtqylweryfgsiekivftfeller--lpdyprvssly----- 441
Qy      559 KAAGTDPNMADNGDDLGPVETEPYASPRYLRLNFTLVTDYFRSMNCAPGPFHLFP-HT 617

```

```

Db 442 -----wmspgallltpft 456
Oy 618 PFDPLVPEAKFLGSGTGCFRPIGGAGSGCKEFOALGCVNPREQTVMLTYREEVLM 677
Db 457 dsxymp-----flnsmgaeprrhnytaivlyqgaalgrna-----aly 496
Oy 678 NSLERLNLCPILKKNVVVWMSPK-LSEEDLMPDIGVPIWV-----RT 720
Db 497 klvtllksqfverellvwaedrplrlkktwprshrlphvlsjgsrtsgaagpqsqt 556
Oy 721 E-KNSLNRPFLPMWIEIEATLSDDDAHLRHDEIMGFVFWREARORTVGFPGGRHW 779
Db 557 egrpslgrtlrpyedldavlsidedallntcedlftayltwrdfrerlygypratanfwd 616
Oy 780 IPHOSMLTNSNYSCELSKVLTLGAFAFPFKRYAVALTSYVMPAIDMVDXYINCEDIAMN 839
Db 617 daknawgyltkskwtlnyalsvltlgaafylnrynylvtmvsllllktvgqsncedllmnl 676
Oy 840 VSHITRKRPPIKVTYSRMWFRDCGCQALSHDOSHHEHNRKACINFPVKYGYGPIRLLYQFRY 899
Db 677 vshvtrkrlpklvtlvtqtkgkldretqspwdrbdlfvgscntfaavfgympllrlslm 736
Oy 900 dsVLEK 905
Db 737 dpmlyr 742

```

Search completed: September 4, 2002, 14:59:46  
Job time: 166 sec



**THIS PAGE BLANK (USPTO)**